R E P O R T R E S U M E S

ED 003 675 24

THE DEVELOPMENT OF AN INSTRUMENT TO MEASURE CASEWORK KNOWLEDGE, THE TULANE ASSESSMENT SCALE FOR CASEWORKERS.

BY- GOLDSTEIN, HARRIS K.
TULANE UNIV. OF LOUISIANA, NEW ORLEANS

REPORT NUMBER CRP-S-062 PUB DATE JUN 65
EDRS PRICE MF-$0.18 HC-$3.92 98P. TULANE STUDIES IN SOCIAL WELFARE, VOL. VII, 1965

DESCRIPTORS- *MEASUREMENT INSTRUMENTS, SOCIAL WORK, *CASEWORKERS, *SOCIAL WORKERS, STUDENT EVALUATION, *EVALUATION TECHNIQUES, *RATING SCALES, TEST CONSTRUCTION, TEST VALIDITY, TESTING, NEW ORLEANS, LOUISIANA

REPORTED WERE SOME PROBLEMS ENCOUNTERED IN MEASURING SOCIAL WORK COMPETENCE OR SOCIAL CASEWORK KNOWLEDGE, SOME PREVIOUS ATTEMPTS TO SOLVE THESE PROBLEMS, AND THE DEVELOPMENT OF THE "TULANE ASSESSMENT SCALE" (TASC) FOR CASEWORKERS. TASC WAS ADMINISTERED TO A TOTAL OF 378 RESPONDENTS, INCLUDING FIRST- AND SECOND-YEAR STUDENTS FROM SCHOOLS OF SOCIAL WORK AND PRACTICING CASEWORKERS WITH MASTER'S DEGREES. ITS PURPOSE WAS TO MEASURE CERTAIN IMPORTANT ASPECTS OF CASEWORK COMPETENCE, EMPHASIZING OBSERVATIONS OF WHAT A STUDENT OR WORKER KNEW, RATHER THAN TO REVIEW THE EXPERIENCES OR THE METHODS THAT HELPED IN HIS PROFESSIONAL DEVELOPMENT. THE RELIABILITY OF THE TASC WAS FOUND TO BE EQUAL TO OR BETTER THAN MANY OTHER INSTRUMENTS IN COMMON USE IN PSYCHOLOGY AND SOCIOLOGY AS WELL AS OTHERS IN USE IN THE FIELD OF SOCIAL WORK. IT WAS CONSIDERED A USEFUL TEST BECAUSE IT COULD BE ADMINISTERED IN ONLY 1 1/2 HOURS TO AS MANY AS 100 RESPONDENTS. (JH)

ED003675

S-062

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

S-062

TULANE STUDIES

in

SOCIAL WELFARE

The Development of an Instrument to Measure Social Casework Knowledge

(The Tulane Assessment Scale for Caseworkers)

School of Social Work

Tulane University

Vol. VII, New Orleans, La., 1965

ED003675

THE DEVELOPMENT OF AN INSTRUMENT TO MEASURE CASEWORK KNOWLEDGE

(The Tulane Assessment Scale for Caseworkers)

Harris K. Goldstein, D.S.W.

Professor of Social Work

TULANE UNIVERSITY

June 1965

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGEMENTS	i
Introduction	1
Chapter I SOCIAL WORK COMPETENCE AND SOCIAL WORK KNOWLEDGE	5
Chapter II PROBLEMS AND ATTEMPTED SOLUTIONS IN MEASURING CASEWORK KNOWLEDGE	15
Chapter III THE MEASURING INSTRUMENT	23
Chapter IV FINDINGS	44
Chapter V SUMMARY AND FUTURE POSSIBILITIES	64
APPENDIX	70

ACKNOWLEDGMENTS

This instrument has been called the Tulane Scale because so many people at Tulane helped with its development. Persons at other schools, however, also assisted. Faculty at the George Warren Brown School of Social Work, Washington University, were active during its early stages. The final development could not have been completed without the excellent help and cooperation of the Denver and Tennessee Schools of Social Work.

The idea for this instrument was originated at Tulane by a faculty committee composed of Mary Lewis, Dorothy Howerton, and Walter Kindelsperger. Many of the early questions were developed by Dorriece Pirtle and Mary Schulte at Washington University. Questions for later forms of the instrument were developed by Mary Lewis, Rita Comarda, Helen Cassidy and the writer at Tulane.

At Denver and Tennessee, James Kern and Willie Bratton administered the instrument to caseworker and student groups tested in Denver and Nashville. They provided suggestions for improving the administration of the Scale and for obtaining data on students' grades and workers' rating scales.

Edward Cryer, Terry Faget, Helene Hartmann, Laura McCall and Claudia Odom who served as professional judges did considerably more than merely score responses. Their reactions to the instructions for judges were most helpful in getting these into the most understandable

form. The clerical judges, Frances Gautreaux and Eunice Dannemann, also assisted in reaching this goal.

The secretarial staff at Tulane was most helpful with the format of the manuscript and is due credit for its general appearance.

This research was supported through the Cooperative Research Program of the Office of Education, U.S. Department of Health, Education and Welfare.

THE DEVELOPMENT OF AN INSTRUMENT TO MEASURE CASEWORK KNOWLEDGE

(The Tulane Assessment Scale for Caseworkers)

Harris K. Goldstein, D.S.W.

Introduction

Like other professions, social work has tried to insure the competence of its members; like other professions also, it has generally assumed that competence was achieved if its members had certain educational and practice experiences. Social work does not have, at this point, any objective measures of competence that can produce consistent findings when applied during or at the end of the educational process. The development of such a measure for casework knowledge was the object of the study reported in this material.

The continued emphasis, even in recent times, that competence for graduating students is reached by going through a designated method of education can be seen in a statement by the Council of Social Work Education:

> In order that the student maintain a level of competence necessary for responsible entry into professional practice, and sufficient to serve as a basis for continuing professional development, concentrations are offered.[1]

---

[1] Official curriculum policy for MSW degree program in graduate professional schools of social work, p. 11. Approved by the Council of Social Work Education Board of Directors, October 19, 1962.

When the profession extended its concern to certifying the competence of workers, it placed a similar emphasis on measuring competence methodologically. Two years under the supervision of an NASW member was the basis for certification as an ACSW, rather than an assurance that the worker had a certain amount of knowledge or could perform certain tasks.

Minimum standards for private practice were also set by the 1962 Delegate Assembly in terms of experience rather than by any measure of what a worker knew or could do. A report of this meeting mentions:

> . . . five years of acceptable full time experience in agencies providing supervision by professionally trained workers, of which two years were in one agency consecutively under such supervision, while giving direct service and using the method or methods to be used in the practice contemplated.[2]

While the 1964 report from the NASW Committee on Competence suggests that objective measures of competence be developed, the report continues to suggest interim use of measures like progress in the level of responsibility, continuation of employment, duration of individual in jobs, etc. All of these measures make use of the experience of the person for purposes of evaluation, rather than what the person knows or what he can do at a given time.

Questions occur in day by day educational and practice activities that appear to require a measure of what the student or worker knows or can do at a particular time rather than an evaluation of experiences he has been through. Some of these questions are as follows: How does one know what is the best method of teaching in schools of social work? To what kind of tasks should first and second year students be assigned?

---

[2]Delegate Assembly NASW, 1962.

At what level of competence can beginning workers be expected to perform? On what criteria should retention, promotion, and reassignment of staff be made?

The development of the instrument reported on in this paper was stimulated by some of these practical questions. This instrument attempted to measure certain generally considered important aspects of competence in casework, but it differs from previous attempts in an important way. It emphasizes observations of what the student or worker knows and can do at a particular time, rather than a review of what experience he or she has been through, or the methods that have been used in helping him develop professionally. Specifically, it measures what the student and worker knows about selecting appropriate goals in social casework situations and what the student or worker would propose to do to reach these goals in specified instances. It asks directly of the student or worker being tested what goals he or she would set in certain situations, rather than inferring that the appropriate goals are known because the student or worker has certain educational training or experience. It notes directly whether students and workers can cite appropriate methods of reaching these designated goals in certain practice situations, instead of inferring that if certain experiences or education have been obtained the student or worker will be able to do this. Because the instrument referred to in this study has been under development since 1961,[3] it can be said to some extent to anticipate the decision of the Committee on

---

[3]Tulane Studies in Social Work, Vol. IV, "An Experimental Design for First Year Field Instruction," Mary Lewis, Dorothy K. Howerton, and Walter Kindelsperger, 1961.

Competence in 1964 which asked for "an experimental objective test for measuring competence."[4]

This report covers some of the problems inevitably encountered in measuring social work competence or social casework knowledge. It discusses some of the previous attempts to solve these problems and then describes the development of the present instrument. The compromises made in this development and details on how the instrument was used in this particular study are presented. The fourth chapter presents the findings from administering this instrument to a total of 378 respondents, including first and second year students from the Denver, Tennessee and Tulane Schools of Social Work and practicing caseworkers with Master's degrees from Denver, Nashville and New Orleans. The final chapter offers a summary of results obtained and some suggestions for further research.

A manual for using the Tulane Assessment Scale for Caseworkers is available separately to accredited schools of social work and social agencies. This contains the questions used on the test, appropriate answers to be used in judging responses, instructions for administering the test, training judges, etc. A promise to keep this material confidential within the agency or school to whom this material is sent is required before this material will be released.

---

[4]Report of the Committee on the Study of Competence of NASW Delegate Assembly, November 1, 1964, mimeographed, p. 7.

Chapter I

SOCIAL WORK COMPETENCE AND SOCIAL WORK KNOWLEDGE

Social Work Knowledge as a Part of Competence:

The lack of an agreed on conceptualization or definition of competence has probably been one reason that competence has been certified on the basis of certain easily designated experiences; this lack of a definition continues to cause a problem when one wishes to use observations of current behavior as objective evidence of competence. Not only is there no agreed on conceptualization or definition of what is a competent social worker, or how he or she is to behave if he or she is to be considered competent, but there is the same lack of agreement in the various methodological specialities which make up social work. Though one might expect more progress to have been made in defining competence in casework, on the grounds that there are more persons available to work on the problems, no definition of casework competence has emerged. Though one might expect more activity from researchers, on the basis they are interested in clarifying concepts and definitions, they too have not developed any definition of research competence.

There are further questions whether competence should be conceptualized as a discrete or continuous concept and whether it is to be spoken of in the abstract or only in terms of what one is competent for. Some see competence as dichotomized into competence and incompetence. Others see it measured on a continuum of degrees of competence. If the latter decision is made, a point on the continuum must be selected above which persons will

be considered satisfactory for employment in the profession and below which they will not be permitted to practice. Competence in the abstract appears to be an unmanageable term. It can only be studied as one speaks of competence for something, or in the broad sense, competence for social work practice.

The Commission on Practice recognized that social work practice must first be defined before competence can be studied. In 1958 this Commission published a "working definition of social work practice" which stated that social work practice, like the practice of other professions is recognized as a constellation of values, purpose, sanction, knowledge, and method."[5]

The emphasis in this paper on measuring the knowledge aspects of social casework competence in no way implies a lesser importance to the value, purpose, sanction, and method components of practice. Values were not studied because they were considered to be an influence on how competence or knowledge was used rather than an aspect of competence itself. While sanction and purpose have to do with society permitting social work to operate and with social work's own statement of purpose, these aspects of practice were considered to be part of the setting in which competence is measured rather than a part of competence to be measured. The method component of practice was considered closely related to the knowledge component and to some extent is considered to be measured by the instrument reported on.

Development of ways of studying all of these other variables, values, purpose, sanction, and method, also seems necessary if social work is to be more specific about its practice; however, the core of social work competence can be considered to be its knowledge base, and the professional's

---

[5] Harriett M. Bartlett, "Toward Clarification and Improvement of Social Work Practice," Social Work, April 1958, pages 3-9.

ability to use this knowledge in practice situations. Thus this study focused on the measurement of the knowledge aspects of competence and to some extent the skill represented by the use of knowledge. Its focus on casework knowledge and the use of casework knowledge, rather than social work knowledge, is largely due to the interests of the writer and others at Tulane.

Previous Attempts to Conceptualize Casework Knowledge:

Focusing on the limited aspects of social casework competence called social casework knowledge does not, however, eliminate either the problem of what is meant by this kind of knowledge or what topics or variables are to be studied to determine how much knowledge is present.

This study uses the definition of knowledge proposed by William Gordon. "Knowledge designates generalized perceptions of man and his world which can be symbolized explicitly enough to be reliably communicated, and are susceptible to testing and extension by the procedures of empirical sciences."[6] This definition contains what are considered to be the important aspects of knowledge, the method of obtaining it, by a combination of empirical (sensory) and theoretical (rational) means, the necessity for communicating it to others consistently, and its testability by observation.

Social workers have spent considerable time and effort trying to conceptualize what are the "generalized perceptions which can be symbolized" that should be used in studying what knowledge it is that caseworkers have. Attempts to provide categories for use in assessing knowledge

---

[6] William E. Gordon, "A Critique of the Working Definition," Social Work, October, 1962, pp. 4-13.

and performance can be found in papers on both the selection of workers and their evaluation on the job. As the profession has developed, these attempts have moved from generalized categories to specific ones, and from emphasis on personal qualities to behavioral statements. There is still considerable lack of agreement on definitions of abstract terms, or what behaviors are to be observed.

In 1930, Mary Richmond spoke of "learning to apply reasonable theory to many concrete situations and modifying the theory by results."[7] Other examples of early global conceptualizations are found in Kaufman and Malmud's 1940 "Evaluation of Staff Members in Private Family Agencies" and Hanford's paper from the 1942 National Conference of Social Workers.

Kaufman and Malmud include such topics as "how does she relate?", "ability to understand the problem presented by the client;" "dynamic use of understanding of the client situation in treatment and ability to carry out treatment plans, etc."[8]

Hanford's paper, in the National Conference of Social Work Proceedings for the year 1942, sets the qualifications for a worker for a private family agency as including, besides graduation from a school of social work, "enough security in practice to insure ability to carry a range of cases; enough maturity to organize one's work, accept responsibility, etc.; and personality free from gross difficulties."[9]

---

[7] Mary Richmond, "The Long View, Papers and Addresses," ed. by Joanna Colcord, 1930, p. 95. Russell Sage Foundation.

[8] Margaret Kaufman and Helen Malmud, "Evaluation of Staff Members in Private Family Agencies," Family Service Association of America, 1940.

[9] Jeannette Hanford, "Standards of Measurement Used in Staff Evaluation in a Private Family Agency," National Conference of Social Work, 1952, pp. 532-538.

A few years later one notes more emphasis on performance and a beginning acknowledgement of the necessity for defining abstract terms in concrete ways. McCaffery's paper is an example of this and indicates not only some of the variables on which knowledge and skill are to be evaluated but also some evidence of how much of each is expected.[10] She suggests such topics as, "skills in interviewing, some new conceptions of behavior, a beginning understanding of self-control, consciously aware of different levels of casework treatment," etc. These deal to some extent with qualities of the worker as much as they do with performance, and are at a higher level of generality than many would want today.

In 1950, Chichester speaks of second-year students in terms of various "abilities" they should have, such as the ability to identify with the client on an objective basis, the ability to sustain a relationship, etc. The way some of the categories are described in this particular paper illustrates another problem in the measurement of knowledge, that of setting the standards for it. Many of the statements here use the terms "more" or "movement toward" indicating that a gain is expected, but there is little mention of exactly where the student should be at a particular time.[11]

More recently, topics on which evaluation is to be based have been made more specific and concrete, but there are still unsupported implications that everyone knows the meaning of certain terms. Schubert uses a number of specific topics or variables but there are still some terms that

---

[10] Miriam McCaffery, "Criteria for Student Progress in Field Work," Journal of Social Casework, January, 1947.

[11] Elizabeth Chichester, et al., "Field Work Criteria for Second Year Casework Students," Principles and Techniques in Social Casework 1940-1950, ed. Cora Cassius, (FSAA, 1953).

are not defined.[12] Perhaps these definitions were omitted on the assumption that there is a general understanding of what is meant, but there is some evidence that such an assumption is unwarranted. Included in Schubert's study are such variables as "worker's feelings toward the client, worker's perceptions of social and economic needs of the client, understanding of the relationship," etc. That further definition of at least some of these terms may be needed is shown by a study completed, as was Schubert's, as a doctoral dissertation. Shively found almost no agreement on the meaning of even such a cornerstone term as the "social work relationship."[13]

That there is still considerable difference of opinion about what variables or topics should be used as a basis for evaluation is shown by a recent study specifically about this difference of opinion. Conner, in 1961, found that approximately one hundred supervisors agreed on only about one-half of certain variables presented to them when they were asked which of these were essential in evaluating performance (67 out of 129 variables). Even more significantly, she found little intrasubjective agreement when the same supervisors were asked on two occasions to select these variables.[14]

---

[12] Margaret Schubert, "Field Work Performance: Achievement Levels of First Year Students in Selected Aspects of Casework Service," Social Service Review, Vol. 32, pp. 120-137.

[13] Frances Shively, "A Study of the Casework Relationship," (unpublished doctoral dissertation, George Warren Brown School of Social Work, Washington University, St. Louis, Missouri, 1962).

[14] Geraldine Conner, "A Study of Indices of Competence Used by Supervisors to Evaluate Caseworker Performance," (unpublished doctoral dissertation, George Warren Brown School of Social Work, Washington University, St. Louis, Missouri, 1962).

Criteria for Studying Knowledge Based on Current Behavior:

Any current attempt to measure casework knowledge on the basis of behavior at a particular time, rather than on the basis of the experience through which the caseworker has gone, must meet certain criteria. There are criteria for the topics or variables selected; for the data sources from which evidence is taken; for the methods used to gather evidence and to note qualitative or quantitative differences from person to person; and for the level of knowledge that is to be considered adequate or satisfactory.

The first criteria for topics or variables is that they adequately cover the area to be measured. Practicality, however, usually requires that the topics sample rather than cover every area of knowledge. This covering or sampling may be achieved in two ways: by some previously developed or theoretical formulation of what casework knowledge covers; or by some inductive analysis of sample observations made in practice. Until a theory of casework is better developed, many would feel that the latter approach is more likely to assure that the variables selected are representative of knowledge required for current practice.

Whatever topics or variables are selected as relevant to measuring casework knowledge, each must be clearly conceptualized. Abstract terms must be explicated in such a way that it is clear how evidence for each can be found. Each topic should be as discrete from the others and as mutually exclusive as possible. Independent observations for each may then be made and more valid and objective measurements are thus possible.

Whether topics or variables are selected on a theoretical basis or by observing practice, the data source should be a primary one; further, the data source should provide some kind of discrete units for observation.

The case record usually used as a basis for measuring knowledge is not a primary data source, since what has occurred in the interview is screened through what the worker has decided to record. Actual observation of an interview makes use of primary, rather than this secondary kind of data. Without specification of what units are to be observed, however, an interview provides too many possible alternate units. Specification must be made whether words, gestures, affect, sentences, etc. are the units, so all can observe the same thing. Any method of observation proposed must assist a number of observers to be able to consistently and validly report the same findings when observing the same data. Further, the method of observation should permit discrimination of one group from another, if it does not discriminate between individuals.

Researchers use the term "reliability" of measurements or observations to indicate that there is agreement or consistency among different observers. If different observers agree on the same observation, one can say that the observations are reliable and this is generally considered to be at least partial evidence they are objective and relatively free from personal bias. The term "reliability" will be used in this sense in this paper. No definite standards have been set for how much agreement one should have, but generally there is an expectation that agreement or reliability between two different observers should be as good as that obtained in psychology or sociology in well carried out research studies. This is generally taken to mean that judges should agree on 80% or better of their observations. Put in statistical terms, this means that there should be a correlation between observers of .70 or higher when both make the same observations.

The other criteria for observations is frequently given the name "validity." Validity is the measure of whether observations are correct,

that is, whether they measure what one intended to measure or whether observations agree with some previously agreed on criteria. When casework knowledge is being measured, one can say the measurement is valid when it agrees with other previously acceptable measures of casework knowledge, such as supervisory judgments or grades in school.

Discrimination of measurements or observations are often considered satisfactory when the method used discriminates one relatively homogeneous group from another. Discrimination of one individual within a homogeneous group from another individual within this group is, of course, more desirable but more often not achieved. Lack of discrimination between individuals is most likely due to the fact that measuring methods are not free from error. The relative knowledge of two individuals may be so close that human judgment cannot differentiate between them. This situation may be likened to a situation where one is measuring with a ruler that has marks on it only at the foot and half-foot intervals. The ruler will differentiate groups who differ in height by six inches, though it will not differentiate people who are different by an inch or two.

Criteria for the level of knowledge considered acceptable depend partly on the expectation by society of social work and partly on social work's expectations of itself. The consideration of competence as a continuum (as was done in this study), means that the knowledge on which competence is based must also be considered a continuum. Thus many different levels of knowledge are possible. Expectations of knowledge held may vary considerably for persons with different levels of training and experience who are expected to carry out different tasks. At some point the profession must agree what is a reasonable level of knowledge to expect,

either the minimum amount for certification or for each task. This expectancy must be set on each of the variables or topics that are used to measure knowledge.

Currently, the profession is still working on this problem, and this study does not attempt to make any statements about what level of knowledge is acceptable. The norms and ranges of knowledge obtained in this study, however, may provide a basis for decisions about these standards. By further study of the relationship between outcomes for clients and knowledge shown by their workers, social work may be able to decide what knowledge is necessary to reach certain desirable outcomes. There are already a number of studies of the outcome of social work services;[15] later studies may also measure this relationship.

---

[15] A number of studies have measured knowledge indirectly by studying what has happened to clients. When clients improved it was assumed workers had more knowledge and when they did not, it was assumed that workers had less. Some of the studies which have attempted to evaluate knowledge indirectly by studying influences upon clients are Henry M. Graham and David Blumenthal, "Why We Fail -- As Clients See It," Family Service Highlights, 1955, pp. 91-94; Robert G. Ballard and Emily H. Mudd, "On Theoretical and Practical Problems in Evaluating Effectiveness of Counseling," Social Casework, Vol. 38, 1957, pp. 533-538; Ann W. Shine, "Evaluation of Results in Social Work," Social Work, Vol. 8, Oct. 1963, pp. 26-38.

Chapter II

PROBLEMS AND ATTEMPTED SOLUTIONS IN MEASURING CASEWORK KNOWLEDGE

An instrument to measure casework knowledge obviously should try to meet all the criteria mentioned in the previous chapter. In this study, because of the limited agreement in the field regarding what variables should be measured, emphasis was placed on development of a method that could be used with any combination of variables. The validity or correctness of variables selected was assumed satisfactory if findings agreed with other measures of casework knowledge. This chapter summarized the major measurement problems in measuring casework knowledge and discussed various attempts to eliminate or reduce these problems.

The difficulties in obtaining reliable, valid and discriminating measures of social work knowledge may be put into one of four categories.[16] There are problems in providing a standardized situation, problems in planning for a standardized response, problems in setting levels of acceptable knowledge, and problems introduced by extraneous variables and/or interaction between these three areas.

Problems in Providing a Standard Stimulus:

First, any situation serving as a basis for measuring knowledge must be one that all persons whose knowledge is being measured can experience identically. Unless it is this kind of situation, the measurement

---

[16] Alfred Kadushin, "Testing Diagnostic Competence -- A Problem for Research," (mimeod material, April 1961) has done a thorough job of reporting many of these difficulties that relate to diagnosis. The references that follow have drawn heavily on his paper.

is influenced by the situation as well as by what the person being tested knows. Ideally, this problem would be solved if all students or caseworkers whose knowledge was being measured had a chance to work on exactly the same case at the same time, but to do so independently of each other. If clients were duplicated as in the "Brave New World" of Huxley, one could have two, thirty-two or even a larger number of identical clients. Without such a plan, current methods attempt to find "similar" cases; while this method may be useful for gross measurements, it has obvious limitations for refined comparisons.

In the framework of contemporary psychology, this problem is formulated by saying everyone whose knowledge is being tested must have a chance to react to a standard stimulus. At present, this can be achieved only by simulating a case situation or giving the respondents a vicarious experience of some kind.

Problems in Providing Standardized Responses:

Besides the standard stimulus, or a standardized situation to which caseworkers or students must react, there must be standardized ways of responding to it. From an ideal standpoint this problem would be solved if all respondents possessed a series of stereotyped responses, one of which they could select in the measurement situation. Traditional methods of measuring casework knowledge, of course, do not use standardized responses. Not only are caseworkers likely to respond differently to the same situation (and from the standpoint of helping the client this may be beneficial) but the records of their responses which are used as a source of data may often vary further. Different agencies may require more detail in recording than other agencies or give emphasis to some points at the expense

of others. A compromise between asking respondents to use one of a series of stereotyped responses and permitting any kind of individualized responses is to set limits on the variation permitted. One way to limit variation is to limit the time permitted for the response (which incidentally is a realistic simulation of practice).

Problems in Setting Acceptable Levels of Responses:

A third source of difficulty is how to standardize the method used to set what is an acceptable level of knowledge. This is necessary so that standards can be duplicated elsewhere. In the traditional method where the supervisor's judgment is used to decide whether the caseworker has done well enough, there is obviously considerable possibility for difference in standards from supervisor to supervisor. How much knowledge a person holds then becomes a function of the person making the judgment. While this problem cannot be eliminated completely any more than the other problems above, it can be handled by giving all judges an agreed-on standard to use. The more explicit and clear this standard is, the less this particular problem occurs.

Problems in Reducing Extraneous Variables:

A fourth source of difficulty is the introduction of uncontrolled variables into the situation by interaction of some of the above problems. Where the measuring instrument is a field instructor or supervisor, the judged level of adequacy may be dependent on other points than the stimulus itself. In practice the field instructor may make allowances for the fact that the student had a harder-than-average case, or showed an ingenious response to it, etc. Because supervisors almost always know more

about their supervisees than is shown in any one case to which the supervisee has responded, this is a particular danger in the traditional methods of measuring knowledge. It can be best avoided in any simulated situation by making sure the person making judgments does not know the person being judged.

There have been many attempts to eliminate one or more of these problems, all of which have been successful in various degrees.

Attempts to Find a Standard Situation or Stimulus:

Developing a standard situation to which students or caseworkers can react requires a choice between two mutually contradictory alternatives. The more standardized the situation, the less it is like an actual practice encounter between worker and client; conversely the more realistic the situation, the more difficult it is to standardize it. Listed below are samples (not exhaustive) of some of the attempts to deal with this problem. They are listed in the order that they deviate from the realistic social casework practice situation, with the situations more like practice listed first, and the situations most unlike practice listed last. Conversely the more standardized situations are generally near the end of this list while the least standardized ones are near the beginning.

Use of an actor in the role of a client has been one method used to provide a standardized situation to which a number of respondents may react independently . Here the actor is given certain "lines" to say and certain guides to follow in various situations which may come up. This same actor can be interviewed in turn by a number of students or caseworkers, with observations made of their actions and reactions to the

actor.[17]

Difficulties arise in the use of an actor in this way because of the practice the actor receives in carrying out his role, embellishments the actor may make as he goes along, or other difficulties related to playing the role in the same way. To make the situation realistic, some flexibility must be provided in the actor's "lines" and this makes for difficulties in standardization.

Because of these difficulties and because situations that provide vicarious experiences with the helping process can be kept more standardized, a number of varieties of vicarious situations have been tried. Many of these involve a caseworker and client encounter and the student or worker whose knowledge is being tested is asked to project himself or herself into the situation used as a stimulus and to say what he would do if he was a worker in the situation or to "second-guess" the worker and try to tell why the worker did a certain thing.

Hans Strupp has made use of a motion picture as a vicarious situation to study the personal techniques of psychotherapists.[18] In a study of casework observations, Roger Miller has used a motion picture showing an interview between a psychiatrist and a patient.[19] Donald Brieland used taped records of interviews, combined with transcripts of these tapes in

---

[17] Use of this kind of stimulus has been made in psychology by E. Lowell Kelly and Donald W. Fiske in "The Prediction of Performance in Clinical Psychology," University of Michigan Press, Ann Arbor, Michigan, 1951, and E.J. Thomas and Donna McLeod, et al, "In-Service Training and Reduced Workload, Experiment in a State Department of Welfare," New York, Russell Sage, 1960.

[18] Hans Strupp, "The Effect of Psychotherapist's Personal Analysis Upon His Techniques," Journal of Consulting Psychology, 1955, 19, 197-204.

[19] Roger R. Miller, "An Experimental Study of the Observational Process in Casework," Social Work, 4, 158, 96-102.

a study of how well adoptive workers agreed about prospective adoptive applicants.[20]

Case records or abstracts from case records have been used in a number of studies. One of the more ingenious is that of Carr, who used "bits" from case records, which were arranged in a prescribed order and presented to respondents serially. At various places, the presentation was stopped and respondents were asked to make a diagnosis based on the material they had been given.[21] Projective techniques, including those making use of sentence completion tests have been used in a number of cases. A sentence completion test is currently in use at Washington University, as a part of a study of field instruction, to compare the knowledge of an experimental and control group of students.[22]

Probably the most standardized of all situations (although possibly the most removed from actual practice situations) are the questions asked by a number of State Welfare Agencies in Civil Service Examinations. These present a situation which is described in a sentence or paragraph and offer a number of alternate courses of action. The respondent is asked to select the correct action from among those given.

Attempts to Provide Standardized Responses:

Equally ingenious attempts have been made to provide standardized

---

[20] Donald Brieland, "An Experimental Study of the Selection of Adoptive Parents at Intake," Child Welfare League of America, New York, 1959.

[21] Lewis Carr, "The Relationship Between the Use of Knowledge in Practice and the Effectiveness of Practice as Seen in Psycho-Social Diagnostic Impressions and the Development of Predictions by Social Caseworkers," Doctoral Dissertation, Washington University, St. Louis, Missouri, 1961.

[22] Virginia Turner, "The Happy Congruence of Theoretical and Practical Considerations in Instrument Validation," paper presented at National Conference of Social Welfare, Los Angeles, California, May, 1964.

methods of responding to these various situations. They range from oral responses that have been taped in an effort to protect their fidelity and completeness, through narrative responses in writing, to selecting the correct response from a number of possible responses as on the Civil Service examination as mentioned earlier.

Thomas and Brieland in the studies mentioned earlier have made use of narrative responses and/or selecting an answer from a check list. In psychology, Giedt has made use of rating scales[23] and in social work, Miller has made use of a Q sort.[24]

Attempts to Set Levels of Adequacy:

Both direct and indirect methods have been used to set the criterion for adequate performance. The direct method of setting standards of what knowledge a caseworker should have indicates what kind of behavior should be expected in designated situations. The indirect method sets standards in terms of completing a specific amount of work. The indirect method assumes that if a certain amount of work is completed, the worker has indirectly demonstrated certain knowledge necessary to complete the work.

Among the more common direct methods is the use of expert judgment. Either qualitative or quantitative standards may be developed. The use of grades in schools of social work is one such example of setting a standard by the use of expert judges. The grade set then becomes a consequent standard. Grades are usually based on a single teacher's opinion, but the expert opinion can come from a single expert or from a group. In

---

[23] Harold Giedt, "Comparison of Visual Content and Auditory Cues in Interviewing," Journal of Consulting Psychology, Vol. 19, No. 6, 1955, pp. 407-416.

[24] Miller, op. cit.

the latter case, a group may pool judgments until some kind of compromise or a consensus is reached. Peer performance may also be used in setting standards. This may be based on the average performance of peers or on the performance of superior peers. It is obvious that use of peer performance makes for difficulties in improving performance unless superior peer performance is taken as the acceptable level.

Among the indirect methods have been the standards set by certain agencies that so many interviews of a certain kind should be completed within a designated time interval, or that certain things should happen to clients in various situations, such as closing a case after a certain period.

Chapter 3

THE MEASURING INSTRUMENT

History of the Development of the Present Instrument:

The instrument reported on in this paper was first used in 1960 to "determine the ability to act independently in a new and time-limited situation, that is, to provide as nearly as possible a situation like one in which the student himself would be interviewing a client, and where the student would find it necessary to make an immediate evaluation of the client's responses and/or an immediate decision about his own responses."[25]

While teaching at Washington University, the writer became interested in the potentiality of this instrument. In 1961, with the help of a faculty committee at Washington University, other questions were added to or substituted for the questions originated at Tulane, and the test was given to 29 Washington University students who had completed their first year. The rank order of scores from this administration correlated with students rank order by casework grade.

On the basis of analysis of the scores of this form of the test, analysis of responses to each of the questions, and the relationship of these responses to grades in school, further modifications were made in both the questions asked and in methods of scoring. This third form was

---

[25] Lewis, Howerton and Kindelsperger, op. cit., p. 60.

administered in May 1962 to 36 Tulane first year students and 22 persons who were at that time either Tulane faculty or doctoral students. This form discriminated first year student scores from faculty or doctoral students' scores.[26]

In June, 1962, the writer accepted a teaching position at Tulane, and with the help of a faculty committee at Tulane, further modifications were made in the test. These later modifications were mostly in the development of different methods of scoring the test, with the idea that a different scoring method would increase its reliability and validity. This latter form was given to 35 first year students at Tulane in May, 1963, and with further slight modifications in the method of scoring (based on analysis of these responses) this is the form reported on in this study.[27] In its present form, the instrument has come to be known as the Tulane Assessment Scale for Caseworkers. It seems appropriate that it should carry the name of the school in which it originated. The word formed from the first letters of each of the words in the title both provides a short way of referring to the instrument, and at the same time offers a term that seems fitting to the measurement of knowledge, the TASC.

Methods Used in This Study for Measuring Knowledge:

In research, as in practice, any attempt to maximize one advantage often reduces another advantage or turns it into a disadvantage. The review in the previous chapter of earlier attempts at measuring knowledge

---

[26] See "Two Instruments for Measurement of Social Work Performance," 1962, which can be obtained in mimeographed form from the writer.

[27] See "The Tulane Assessment Scale for Casework Performance,"1963, which can be obtained in mimeographed form from the writer.

indicates the application of this principle. The most ingenious attempts at duplicating practice situations often suffered because of lack of reliability of observations made, questionable validity of findings, problems in discrimination, and difficulties in using and administering the test instrument. Those which provided reliable or valid observations often suffered from some other kind of liability.

The method used in this study was not selected because it solved all of the problems mentioned, but because it appeared to offer suitable compromises on many of them. It is believed that the method used in this study maximizes as many advantages as possible while minimizing disadvantages. The administration of the TASC in its present form requires about one hour and fifteen minutes to one hour and thirty minutes. The length and the fact it can be used for individuals or groups thus makes it appear a practical and reasonable instrument. Respondents report becoming involved in the testing process in a fashion similar to the way they become involved in practice, thus providing some evidence of the realistic nature of the stimulus used. Scoring time for one respondent's answers is 15 minutes on the narrative and one minute on the objective form.

Selection of Variables to be Studied and Stimulus to be Provided Respondents:

The lack of agreement in the field about the variables on which knowledge should be measured led to the decision to avoid, in this study, any selection of variables before the study. Instead, any sample of practice, or a proper simulation of practice, was assumed to offer opportunities for students or caseworkers to demonstrate their knowledge of social casework. Obviously, this assumption required that the sample of practice

or simulation of practice be reasonably representative of practice as a whole. An assumption that almost any sample of practice is representative appears in much of current field instruction, where instructors often assume that a single case offers teaching possibilities for much of the knowledge required for social casework.

Because of some of the difficulties mentioned earlier in getting a standardized piece of practice, a simulation of practice rather than an actual practice experience was used in this study. This simulation had all the difficulties mentioned earlier for any such vicarious experience, but the value of the standardized stimulus was considered to outweigh the difficulties. The particular simulated sample of practice used for the TASC was a thirty-minute motion picture. This film showed a series of casework interviews, between a family caseworker and a client.

A motion picture that combined both visual and auditory stimuli was considered more like actual practice than taped records. Funds and time available did not permit the development of a special motion picture to be used in this study. Even if they had, however, it was difficult in the absence of empirical studies to know on what basis a picture could be made representative of practice, or what was expected of students or workers in the field. Thus this motion picture, based on the experience of a number of persons in the family casework field, was considered representative. Further evidence for its representativeness was considered to be the fact that the film selected was in current use in the field to demonstrate to lay and professional groups the services of family casework agencies.

The likelihood that a simulated experience would be seen by respondents as similar to an actual practice one was originally based on the

subjective reactions of various teachers who viewed the particular film suggested.[28] Respondents, when asked to give their own reactions to the realistic nature of this stimulus, however, supported this assumption. Respondents almost universally reported that they became involved in the situation shown in the motion picture in the same way they did in practice. Further evidences of the reality to respondents of this simulated experience were their involuntary non-verbal and verbal responses as they viewed the motion picture. Not only did facial expressions of the respondents show they were involved, but at tense moments in the film, respondents often involuntarily made tension relieving remarks, while at humorous points in the film, laughter could be heard.

The method used to present this simulated piece of practice to respondents was considered to add further to the realistic nature of the experience, as well as provide further standardization of the stimulus. The motion picture was shown to a group of from 20 to 80 individuals and stopped at various previously determined and standardized points. At each stopping point, the respondent was asked various questions about what was going on in the picture and made responses as indicated below. Stopping the picture at intervals, asking for responses, and then going on with the film was considered more like actual practice than showing a complete series of interviews and asking questions about them as a whole. The method used required the respondent to make an immediate response, much as in actual practice interview. The limited time provided for the response to be made (a maximum of three minutes) required an immediate reaction. Also like practice, each time the film was stopped, the respondent

---

[28] These included Rita Comarda, Helen Cassidy, Dorothy Howerton, and Mary Lewis at Tulane and Dorriece Pirtle and Mary Schulte at Washington University.

had seen what was for him a new situation, one in which he was required to make a quick and immediate decision.

The selection of the twelve stopping places for this motion picture may be considered to be at least a somwehat random sample of the variables on which knowledge of casework should be measured. The selection of stopping places attempted to maximize the questions' representativeness of practice, the discrimination required in answers, and the reliability with which answers could be scored. During the four years of the development of the TASC, a total of 26 different stopping places in the motion picture had been used one time or another. These varied in how well scores for responses discriminated between students and experienced caseworkers who saw the film, and in terms of which responses could be judged most reliably. Some questions also appeared to provide a more varied sample of caseworker activities, and had more activity than others between them in the motion picture to provide continuity and further stimulation to the respondents. In the opinion of the faculty group, the twelve stopping places maximized all these goals; twelve was also the maximum number of interruptions of this thirty-minute film that could be made without causing respondents to lose continuity.

Method of Obtaining Responses:

Two methods of obtaining responses have been used, one for what is referred to as the narrative form of the TASC, and one for what is called the objective form. The method of obtaining responses for the narrative form attempted to provide a natural way of obtaining responses, and at the same time set limits on the time available to respond. Because in practice,

it is customary to write narrative reports of helping activities with clients, a narrative response was asked for. To limit responses to a realistic length, time provided for responses was controlled.

Before the motion picture began, respondents taking the narrative form of the TASC were seated in rows in a room facing the motion picture screen. Prior to starting the film, each respondent was provided with a booklet consisting of alternate blank sheets and a sheet for each of the questions (with space provided for writing in answers), all with an identifying number on each sheet. The respondent did not see the first question until the film had been stopped for the first time. After answering the first question, the respondent was asked to tear off his answer sheet and pass it to the end of the row where it was picked up by a monitor. The respondent was thus prevented from changing responses on the basis of information that might be provided later by the motion picture. The respondent then found on the top of his booklet an alternate blank sheet that prevented him from seeing the next question, while he watched the motion picture again. When the picture was stopped for the next question he tore off the blank sheet, saw his next question, wrote his answer, passed his answer to the end of the row and repeated this procedure until the film ended.

For the objective form, each respondent was seated in a similar fashion, and provided with similar booklets with identical questions. The booklets differed only in that the objective form provided four alternate answers to each question. Directions asked the respondent to select one answer by circling an identifying letter before it.

For testing the comparability of scores on the narrative form and

the objective form, certain groups were given both booklets before the test. At each stopping place, they were asked to complete the answer to the narrative form, turn this in to the person giving the test, and then complete and turn in the answer to the objective form before the picture was started again.

Six types of questions were asked in these booklets. For the type one questions, the motion picture was stopped after something the caseworker in the film had said. The respondent was asked, "What did the worker hope to accomplish or why did the worker say what she did?" For the six questions of type two, the motion picture was stopped after something the client had said. The respondent was asked two questions at each of these points. Separate answers were expected to each question and a separate place on the answer sheet provided for each. The questions were: "What would you do or say at this point? Include non-verbal responses if appropriate," and "Tell what you want to accomplish or tell why you would say or do this." The two type three items asked for diagnoses at the end of two different interviews, one for the wife and one for the husband of the family receiving treatment. The single type four question occurred at the end of an interview and asked about the appropriateness or inappropriateness of a subject change made by the caseworker in the film during this interview. The respondent was also asked for reasons for this decision. The one type five item asked for the method and technique used at a specific point where the film was stopped and for what was accomplished by this technique. A single type six item was an objective item providing four alternative courses of action for handling this case at the place where the motion picture ended. The respondent was asked to select one of these

alternatives as being correct.

The time provided for each response to the narrative form of the TASC was two or three minutes. Less time was needed on the objective form and one minute was allowed. These periods were determined by experiences in giving the test in previous years. Experienced caseworkers sometimes wrote longer answers than students, although there was no significant difference between the length of answers for first and second year students. Experience showed that 90% of the respondents, even when these were experienced caseworkers, found the three minute period adequate to write their answer to any of the questions and for some questions two minutes were adquate. The procedure while administering the test was to note when respondents had finished, and when approximately 95% of the group stopped writing and lifted their heads, they were asked to pass their papers to the end of the row. Usually two-thirds of the group finished within the same thirty second period. Providing this limited time, rather than a longer period was found to prevent conversations between persons who had finished and to reduce confusion in general. Before such a limited time had been set, conversations sometimes started between persons in neighboring seats and these tended to delay the administration of the film and to interfere with others hearing the first words of the film when it was started again.

The presence of a monitor for each 25 respondents was found to be sufficient to permit responses to be picked up in a few seconds, and for the motion picture to be started again with a minimum of delay. When only 25 respondents were present it was found that the person operating the projector could also pick up the responses. However, when more personnel was available it was found that having one person operate the projector

and one person pick up the responses facilitated the actual administration of the test.

Methods of Classifying Responses:

The criterion or standard used for classifying or scoring responses on both the narrative and objective forms of the TASC was based on a consensus of "expert judges." In arriving at this consensus, the Tulane committee discussed each question separately and considered what kind of response could be considered satisfactory for a student who had completed the first year of graduate training. Use was made of responses, obtained in the 1962 and 1963 testing of first year students as typical. Various possibilities were considered until there was general agreement on each question.[29]

The decision to select students who had completed one year of training as the group on whom to standardize answers was based on several considerations. In the first place, the TASC was originally developed to measure knowledge at the end of the student's first year. Second, the end of the first year is midway in the student's school experience. It was assumed that a test that measured knowledge at the end of the first year would not be so easy that all second year students would be scored at the top of the scale, nor so difficult that any students tested at the beginning of their first year would all score together at the bottom of the scale. Experience with the test appeared to bear out these expectations. When responses of first year students in 1963 were scored

---

[29] This committee consisted of Mary Lewis, who had been Director of Field Instruction during the original study of Field Instruction at Tulane, Helen Cassidy, who was Director of Field Instruction at the time of this study, Rita Comarda who was Chairman of the Group Work Sequence at the time of this study, and the writer.

according to standards set, scores ranged from 8 to 16 out of the possible 20 points used at that time, so that possibilities for both higher and lower scores remained. In the 1963 administration a number of experienced caseworkers with Master's degrees were given the test and their scores, too, were less than the maximum 20 points possible.

For each question, written criteria were developed and provided for use of judges in the scoring process so that each response could be classified either as "satisfactory," "innocuous," or "unsatisfactory." In general, responses were classified as satisfactory if they forwarded the helping process in some way; "innocuous" responses were those which did not forward the helping process, although they might not have actually hindered it; answers were labeled "unsatisfactory" if they either hindered the helping process, harmed the client, or focused or directed activity in a way that delayed the solution of the client's problem. Alternate answers offered respondents who took the objective form had been selected from responses obtained in 1962 and 1963 that fit these various classifications. In some cases one of the four responses offered was satisfactory, two innocuous and one unsatisfactory. On other questions, other distributions were provided. Objective responses were classified by a predetermined scoring key, developed similarly to criteria for scoring the narrative form.

The procedure of asking judges to classify narrative responses rather than to give the response a numerical score was used because it was considered that judgments would be less influenced by this plan than if judges were asked to assign numbers to responses. Numbers were later given to the classifications made by judges, so that material could be

handled statistically, but without the judge knowing the numerical value assigned. A satisfactory response was given a weight of five, an innocuous response a weight of two, and an unsatisfactory response a weight of zero. This decision was made by the writer as most satisfactory after trials of various scoring systems. This weighting was also considered most in keeping with the meaning of the categories above.

The Judging Process:

Judgments of responses were made by four professional judges and two clerical judges.

The four professional judges were chosen to represent various fields of practice and various types of positions.[30] All had MSW degrees and each had more than five years of practice experience plus two years of supervisory experience.[31] Practitioner experience of judges ranged from 5 years to 11 years; their supervisory experience ranged from 2 years to 11 years. Two judges were field instructors employed by the Tulane School of Social Work. One of these had participated to some extent in the development of the study and the other had not. The former had a field instruction unit in the Department of Public Welfare. The latter had a unit in the public school system. One of the other judges was the supervisor in a private family service agency, supervising students from the school and caseworkers from the agency. The fourth judge

---

[30] Because one judge resigned after completing about one-third of her judgments, a total of five judges was used. However, the replacement secured for the missing judge was a type similar in experience, field of practice and type of position so this report is written as though only one person occupied these two positions.

[31] Two were graduates of Tulane; one was from L.S.U. and the other had a Master's degree from New York.

was an experienced medical social caseworker who was not supervising either students or workers at the time of the study but who had done so in the past and who had a responsible position teaching senior medical students and Master of Public Health Students, at the Tulane Medical School. None of the latter three judges mentioned had participated in the development of the TASC. The judge in the private family agency was male; the others were female. Two judges were married and two were single. One judge had seen the motion picture twice; two had seen it once; one had not seen it at all.

The two clerical judges were chosen to represent typical clerical personnel who might be found in most social agencies or schools of social work. The use of clerical judges was an attempt to see if these persons could classify responses similarly to professional judges. If so, then judging in later studies could be carried out at less expense. Both clerical judges had completed high school, but had no other education or training. One had 22 years of experience in a Department of Public Welfare; the other had 10 years as a clerical worker in general and one year in a school of social work. At the time of the study, the clerical duties of both clerical judges included such tasks as transcription of machine dictation, filing and general clerical activities.

Before being asked to classify the responses of students and workers obtained in this study, three of the four professional judges were given approximately three hours and the other judge one hour of training in the judging procedure. This difference in training was introduced deliberately so its influence could be studied. This training was carried out as follows: The three judges receiving three hours of training were

asked to study the criteria for judging responses. After having a chance to read and study these, a group discussion lasting about an hour was held to clarify any questions about the wording of the instructions. There appeared few questions at this point. Judges were then given 15 responses obtained in the 1962 study for each of the five types of questions from the TASC. These particular responses were selected so that variations in responses to each question were included. Each judge then made judgments independently, on this sample of 15 responses, using about one to one and a half hours to do this. The classifications of responses made by various judges were then compared, and in another group meeting of an hour, differences in interpreting instructions were discussed. The fourth judge began her training at this point. Each judge was then given 15 responses obtained in the 1963 study on each of the five types of items and judged these. Interjudge comparisons of categories into which responses were put indicated agreements ranging from 70% to 100%. Mutual understanding of instructions thus appeared sufficiently uniform so judging could begin. The two clerical judges were similarly trained but had only one group meeting after studying the instructions for judging and only one experience in judging sample responses before being asked to judge the responses analyzed in this study.

In both the training period, and in the judging of responses that followed, judges classified all the responses to a particular question before they moved on to another question, rather than classifying all the responses of a single respondent. This procedure was considered to minimize the likelihood that judgment on one question for any particular respondent would influence judgments on other questions answered by the same

respondent. Responses given to judges had been typed to minimize any problems in reading handwriting, and were identified only by a numerical designation given the respondent. All answers to a question followed each other in this numerical order and after all responses to one question were completed, responses to the next question followed in numerical order. Responses of first year students, second year students and caseworkers in the sample judged were interspersed in an order unknown to judges so they did not know the level of education or experience or respondents whose replies they were judging.

Handling of Possible Limitations of the Study:

Some of the limitations of this study arise out of decisions that were made in the course of the study, such as the decision to measure knowledge rather than competence. Others are related to the limitations of the stimulus (motion picture) selected as a sample of practice, limitations due to respondent motivations to participate, and limitations related to judging responses.

One of the most obvious limitations arising out of the decision to measure knowledge rather than performance is that the TASC measures only whether a respondent knows how to help a client. It does not necessarily indicate whether the student or caseworker can do what he says should be done. This is a limitation of any vicarious or simulated experience and not characteristic of this test alone. The profession usually assumes that a student cannot use more than he knows, (although he may and often does use less) and thus the knowledge assessed by the TASC can be considered to represent the upper limits of what the respondent can do. It is further assumed that on the average there is a generalized

relationship between the amount of knowledge that each student and worker has and the amount of knowledge each uses in practice. Thus the TASC can be considered to measure indirectly the relative amount of knowledge that is likely to be used in practice by students and workers.

To make this transfer from measuring knowledge to measuring use, one must know if knowledge and not something else is being measured. Here the definition of knowledge used earlier is helpful. What is being measured by this study <u>does</u> appear to be what the respondent took in through his sensory organs, acted on with his cognitive abilities and communicated to others. The question whether the TASC measures <u>casework</u> knowledge was studied by determining the correspondence of scores in the TASC with other generally agreed on measures of casework knowledge. One such measure for students was their grades in casework courses and field instruction. A similar measure for workers was rating scales completed by the workers' supervisors; in addition, differences in scores between students and workers were compared with differences in training and experience among the respondents.

It was assumed that grades in casework courses and field instruction would be the best measure of casework knowledge available for students and that rating scales completed by supervisors would be the best measures of this knowledge held by caseworkers. Rating scales were selected for workers rather than the usual narrative evaluations so that comparative data could be obtained from different supervisors in different agencies, so that time required for supervisors to obtain this data would be reduced, and so that the administrative task of getting data would be facilitated. There was the further expectation that more objective data would

be obtained from rating scales than from narrative evaluations.

While the use of grades is generally accepted as a measure of casework knowledge, the use of rating scales leaves a question as to whether a newly constructed rating scale actually measures casework knowledge or something else. The assurance that the particular rating scales used in this study measure casework knowledge rests on several methodological steps.

First, the working definition of social work practice was used for the content of the rating scale.[32] Of the nine areas of knowledge mentioned in this definition of practice, six were selected. Knowledge of group process and influences of the group and individuals on each other; knowledge of the community, its internal processes, social services, resources; and knowledge of the social service system, its structure and organization, etc., were omitted. These three kinds of knowledge were eliminated on the grounds that this kind of knowledge was not tested by responses to the motion picture used in this study.

Instructions to raters further attempted to assure that knowledge was being measured by asking the case supervisors to concentrate on the knowledge of each of their workers of attainable goals and the knowledge of methods and techniques of reaching these goals held and used by each caseworker.

Experiences indicated that it might be difficult for case supervisors to differentiate between what the worker knew and how well the worker used this knowledge. A further methodological step attempted to reduce

---

[32] Harriet M. Bartlett, "Toward Clarification and Improvement of Social Work Practice," Social Work, April 1958, pp. 3-9.

this limitation by asking supervisors to rate two scales, one a scale of knowledge held and another a scale of knowledge used. It was assumed that permitting the supervisors to rate a scale of knowledge used would help them to differentiate between knowledge held and knowledge used.

A number of other common techniques generally considered to improve rating scales was used in further attempts to make this scale valid. The number of categories on which raters were asked to rate was limited to four. The even number of categories is considered to reduce what is commonly called "the error of central tendency" or the likelihood that supervisors who had difficulty in making a decision would rate at the center of the scale. Because no absolute measure of casework knowledge was known, a relative rather than an absolute measure of knowledge was asked for. For example, supervisors were asked to rate their workers in group one "if you would place this worker in the top 25% of caseworkers you have known with from three to five years experience;" or the supervisor was asked to rate the worker in group 2 "if you would place this worker below the top 25% but above 50% of the workers you have known with from three to five years experience."[33]

The selection of grades and rating scales as measures of knowledge recognized that both had certain built-in limitations. The reliability of grades, or the percentage of instances where the student would get the same grade again if the grading process was repeated, is unknown. There

---

[33] This scale was originally designed to be used only with caseworkers with from three to five years experience. However, because not enough caseworkers with this limited range of experience were available for the sample required in this study, caseworkers with more or less experience were included. In this case, workers with three to five years of experience were still the group used as a reference.

is a general belief that there is considerable subjectivity about grades and that the reliability of grades varies from course to course and from teacher to teacher.

No simple and easy way to determine the reliability of grades was available. The scope of this study did not permit such a technique as asking each teacher to grade each student twice and comparing or averaging these grades. Even in situations where this has been tried there is always the question whether the two attempts to grade the students are completely independent or whether the memory of the teacher of the first grade is likely to influence the second. If memory does influence the second grade, this kind of attempt may give a false sense of security, and a false sense of reliability is achieved. Whatever their limitations, grades in casework courses and field instruction appear to be a readily available measure of casework knowledge for comparison with TASC scores. The independence of grades of TASC scores was insured by obtaining grades before TASC scores were known.

A previously untested rating scale such as the one used in this study for getting caseworker knowledge has the same liability as grades. Its reliability is unknown. The reliability of rating scales, however, is often assumed to be higher than the reliability of grades. This belief is held on the basis of methodological steps used in constructing rating scales, steps that have been shown on other occasions to bring about desirable reliability. Further assurance of the reliability of rating scales is often based on the fact that there are specific instructions given to the judges to use them.

For the same reasons that no determination of the reliability of grades was made, no attempt was made to determine the reliability of the

rating scale used. Possible influence of memory on any rerating was considered to outweigh any gains made by such a procedure. The rating scale itself was not long enough to permit use of another method sometimes used to determine reliability, such as finding out whether the first half was rated similar to the second half. Since there was only one scale this could not be done.

Differences in the training or experience of respondents were used as the other method of validating that the TASC was an instrument that measured casework knowledge. This was done on the basis that persons with most experience could be expected to have more knowledge and make higher scores while persons with least experience and training, for instance first year students, could be expected to have least knowledge and make lower scores. This was considered true even though the accumulation of knowledge might be made at a gradually diminishing rate as experience increased. In general, it was assumed that the correspondence of differences between groups' scores on the TASC with differences in experience and education of groups tested would validate the instrument as a measure of casework knowledge.

A possible limitation based on the choice of this motion picture is whether findings can be generalized to other situations than those found in family service agency practice. Here, the assumption is held that family service agency practice is one of the more generic kinds of casework services. On this basis, any findings about knowledge held about family casework should be generalized at least to some extent to other fields such as medical and psychiatric social work and child welfare. The validity of this generalization must rest on this assumption until determination can be made if scores of workers from various types of agencies are different. In this study, an agreement was made with participating agencies that no interagency comparisons would be made.

Limitations related to students' or caseworkers' motivations are

unknown, but probably vary from one group to another. There is no way of knowing whether students tried as much as possible to make their responses representative of what knowledge they held. The TASC can thus only be said to measure what casework knowledge respondents reported and not what knowledge respondents had. Again here, however, the assumption is commonly made that there is a relationship between knowledge reported and knowledge held. This assumption is so widespread that it is found throughout the educational process. Grades in school are based on it and degrees awarded on this assumption. Testing conditions probably influenced motivation as well as the performance of the students on the test. At Tulane, for example, when the test was given late in May, the weather was very much hotter than in the other participating cities. At this point, no quantitative determination can be made of the limitations due to motivation, but these must be accepted as present to some degree.

Limitations associated with the judging process are known and can be accounted for since one of the aims of this study was to determine the amount of agreement or disagreement among the various kinds of judges used. The sources of lack of agreement, however, are not known. Some possible sources are differences in the judges' understanding of the responses, differences in the judges' understanding of the instructions for judging the responses, or some differences in the way they saw there two things interacting. Because of the emphasis, through training, on achieving a uniform understanding of the instructions for judging, disagreement appears due to the other two possibilities.

Chapter IV

FINDINGS

The data used in this study were obtained during May, June, and September 1964 through the cooperation of the Denver and Tennessee Schools of Social Work with Tulane. The students tested were in one of these three schools at the time of testing; the practicing caseworkers with Master's degrees were working in the city in which the school was located. In the following material, cooperating schools are numbered rather than referred to by name. The geographical area from which workers came is referred to by the same number given the school. The order in which numbers were assigned has no relation to the order in which schools were discussed previously.

Non-technical discussions of findings for those who lack time and technical background are presented in the body of this chapter; details for those who may wish to use the TASC and compare their findings with those obtained in this study have been provided in the Appendix.

Distribution of Responses by Stage of Professional Development and Type of Training:

A total of 378 respondents provided the data on which findings are based. First year casework students, second year casework students, and practicing caseworkers with Master's degrees made up about equal proportions of this total. The geographical area represented by each of the three schools made up about equal proportions of each of these groups, each considered to represent a stage of professional development. The specific distribution of respondents by stage of professional development and by geographical area is in Table 1.

Table 1

Distribution of Respondents by Stage of Professional Development and Area

Stage of Professional Development	Geographic Location			
	Area 1	Area 2	Area 3	Total
1st Year Students	63	40	49	152
2nd Year Students	37	36	43	116
Caseworkers with MSW degrees	45	37	28	110
	145	113	120	378

In all, the TASC was administered twelve times, once to each of the groups above, once again to each of two groups to get retest scores, and twice to get enough workers from one city. All but one or two of the first year casework students in each of the three schools were included and all second year students at school two and school three were tested. Second year students at school one were volunteers but were found to be a respresentative sample of second year students in terms of the grade point average at this school. First year and second year students at one school were tested at the beginning of the year; students at the other schools were tested at the end of the first year and second year respectively.

Caseworkers who participated were volunteers who responded to invitations to social agencies in the three cities. Since the workers' scores were obtained only to determine if the TASC discriminated between casework students and caseworkers, the fact that the workers were a volunteer sample, and thus not necessarily representative of caseworkers in their city, was not an issue in this study.

Background Factors of Respondents That Were Considered Possible Influences on Scores:

Only those background factors that were considered to influence directly the TASC scores were obtained. These included respondents' years of experience before any graduate training, experience between the two years of graduate training, and after graduate training, and whether they had seen or discussed this motion picture before the study.

Social Work Experience:

There was little difference between caseworkers from the three areas with regard to average experience prior to the MSW degree. Almost all had one year or less. After receiving their MSW, all groups of caseworkers had on the average four to five years of experience; those from one area, however, had significantly more experience on the average than those from the other two areas.[34] Details on the average experience and range of experience of each group of caseworkers are shown in Tables E-1 and E-2 in the Appendix.

Second year students' experience prior to entering school and between the two years was not significantly different from school to school. The average (mean and median) experience for second year students was very close to one year. Details on second year students' experience are in the Appendix in Tables E-3 and E-4. First year students' experience prior to entering school did not vary significantly from school to school; ranges, medians, and means were quite similar to second year students. Details on first year students' experience are in the Appendix in Tables E-5 and E-6.

---

[34] The term significance is used throughout this material to refer to the 5% level.

Previous Experience with the Motion Picture:

Before being involved in this testing experience, more workers than students had seen or discussed the motion picture on which the TASC was based. Even among workers, however, no more than one quarter of any one group had seen the picture before. There was no significant difference from area to area in the percents of each group who had seen or discussed the motion picture before being tested. Details on the percent of each group who had seen and discussed the motion picture before being tested will be found in Tables E-7 and E-8 in the Appendix.

Relation of Background Factors and Scores:

Experience and previously viewing the motion picture before being tested had no significant influence on scores.

In only a very few instances was there a significant correlation between experience of the respondent group and the scores. Students who had some experience before and between the two years made scores quite similar to those students who had no experience and there was no significant difference in the average scores of these two groups. Some experienced workers, however, made lower scores than some workers with less experience (see section on Discrimination - page 53).

Only in two groups of workers was there any significant correlation between scores and having seen the motion picture before. Because so few workers saw the picture before, it appears likely that these correlations were inflated by the kind of distribution tested here. When respondents were divided into those who had seen the motion picture before being tested, and those who had not, there was no significant difference between their scores.

Details on the relationships between background factors and scores are shown in Tables E-9 through E-17 in the Appendix.

Ranges of Average Scores:

Although the TASC discriminated between persons with different levels of professional education (see section on Discrimination), there was considerable overlap in the range of scores obtained by persons at the various stages of development. The ranges of total scores and most frequent range of total scores obtained by each group are shown in Table 2 below.[35]

Table 2

Ranges and Most Frequent Scores by Respondents' Stage of Development

	Beginning 1st. Year	Ending 1st Year-Beginning 2nd Yr.	Ending 2nd Yr.	Workers
Range of Total Scores	35 - 84	36 - 92	40 - 96	43 - 105
Most Frequent Range of Total Scores	60 - 69	70 - 79	80 - 89	80 - 89

Reliability or Consistency of Scores:

Scores on the TASC were found to be stable and consistent over time; different persons who judged the same responses also scored these quite similarly. These findings indicate that questions asked of respondents are relatively clear to them and that respondents do not vary from time to time in how they interpret the meaning of questions; these findings also show that judges are able to interpret instructions for scoring in a similarly consistent fashion.

---

[35] All of the information in this chapter, unless specifically labeled otherwise, refers to the narrative form of the TASC. The objective form is discussed on page 58.

The two groups of students who were tested twice, with an interval of a week in between, first year students from area one, and second year students from area three, made very similar scores on the two occasions. Table 3 below shows the average for the four professional judges for these two groups in both the test and retest situations. There was no significant difference between the test and retest scores as shown in Table 3 below.

Table 3

Test - Retest Scores on Students

	Average Test Scores (means)	Average Retest Scores (means)
Area 1 - first year	79.1	82.5
Area 2 - second year	78.4	77.6

The non-significant difference in test - retest scores and the fact that on one occasion the retest scores were up and on the other occasion the scores went down, would appear to argue that previous experience with the test has little influence on later scores. If found in other test-retest situations later, this is a characteristic of the TASC which can be said to increase its usability. It thus could be used in before and after situations, for example, to determine how much students have learned in courses. Students could be given the test before and after the course. Since these frequencies were obtained on only two groups, however, further testing is necessary before this finding can be considered definitely confirmed. This finding also does not mean that anyone may be permitted to become familiar with the test questions. There was no way to determine in this particular study whether fore-knowledge of the expected responses or study of the questions before being tested had an influence on scores.

Thus the questions and the answers to the questions should be protected and only made available to persons who will score responses.

Correlations between the test - retest scores are shown in Tables R-1 and R-2 in the Appendix.[36]

Similarities in total scores assigned by all six judges scoring the test provide considerable confidence that different persons with different backgrounds and from different parts of the country can score the same responses quite similarly. This finding adds to the potentiality of the TASC being useful in various sections of the country at various schools and by various agencies. The average total scores for each of the nine groups tested showed a very limited variation from judge to judge. The variation within the group of judges was not significantly different from the variation between the student groups at any comparable level of education; variation within the judges' group also was not significantly different from the variation between worker groups.

The average (median) interjudge reliability or agreement on total scores for the 21 items in the test, expressed in correlational terms, was .77. This is comparable to or better than many instruments in common use in the fields of psychology and sociology and comparable to one well known instrument in the field of social work - the Hunt Movement Scale.

None of the factors that had been deliberately introduced for testing had any significant influence on reducing interjudge agreement on total scores. For example, the varied experience of the judges had no relation to how they scored responses. The two professional judges who were

---

[36] Pearsonian correlations were used throughout this study. A number of computations of rank order and phi coefficients were completed on the same data, and no significant differences found when different correlational indices were used.

field instructors agreed about the same as the two professional judges who had come from different social agencies. The agreement among the three professional judges who had seen the motion picture before doing their scoring was almost identical to the agreement between each of these three judges and the one professional judge who had not seen the motion picture before judging scores. The two professional judges who were graduates of Tulane agreed no more closely or widely than the two judges who were not. The judges who were Tulane graduates also agreed about the same with each of the other two judges who were graduates of L. S. U. and New York respectively.

The agreement between non-professional (clerical) judges was not significantly different from the agreement between professional judges; the scores of the non-professional judges were not significantly different from the scores of professional judges. Although the two clerical judges had considerable difference in experience (one year and twenty-two years), there was no significant difference in scores they assigned. Tables R-3 through R-11 in the Appendix show the interjudge agreement for each pair of judges for each of the nine groups tested.

Not only was agreement found on the scoring of the test as a whole but similar close agreement among judges was found on the scoring of the various questions of the test. Expressed in correlational terms, interjudge agreement on questions ranged from .17 to .99. Interjudge agreement on fifty percent of the questions was above .47. The interjudge correlation for certain pairs of judges for each question is shown in Table R-12 in the Appendix.

Discrimination:

In general, the TASC discriminated significantly between the group of first year students at the beginning and groups of first year students

at the end of the first year; it also discriminated between groups of first year students on the one hand and groups of second year students and caseworkers on the other. It did not discriminate between groups of second year students and caseworkers, possibly because some older workers made low scores and some second year students with experience made high scores.[37]

The rank order of scores for the various groups tested was, in general, the same as the rank order of the stages in their professional development. The first year students who were tested at the beginning of their first year scored lowest of all; first year students tested at the end of their first year scored higher; second year students tested at the beginning of their second year were quite similar to the first year students tested at the end of their first year. Highest of all, were second year students tested at the end of their second year, and practicing caseworkers with Master's degrees. These findings are shown in Table 4.

---

[37]Variation within groups of students at the same educational level from the various areas and variation within the group of workers from the various cities was significantly less than the variation between student groups at different educational levels or between worker groups. This kind of analysis showed that the two groups of first year students tested at the end of the first year and the group of second year students tested at the beginning of the second year were sufficiently alike so they could be combined for further analysis. Similarly the two groups of second year students tested at the end of the second year were sufficiently alike to be combined for analysis and the three groups of workers were sufficiently alike to be combined.

Table 4

Average Scores by Respondents' Stage of Development

	Beginning 1st. Yr. Stud.	Ending 1st. Yr. Stud.	Beginning 2nd. Yr. Stud.	Ending 2nd Yr. Stud.	Practicing Caseworkers
Mean*	74.9	81.1	78.4	84.5	84.8
Median Score	68.3-79.0	78.0-85.6	76.0-85.7	78.3-90.5	78.8-91.0

*Mean is the mean of 4 professional judges; median shows range of medians for these 4 judges.

Mean and median scores for all groups are shown in Tables D-1 and D-2 in the Appendix.

Among practicing caseworkers with Master's degrees there was some evidence that relatively long experience (ten to twenty-five years) was associated with low scores. This may have been related to differences in the training these caseworkers received ten to twenty-five years ago, or to the fact that a number of them also had some responsibility for supervision and administration and were somewhat away from casework practice. For example, the fourteen lowest scores made by caseworkers were made by workers with more than ten years experience. Six of these fourteen were engaged to some extent in supervision or administration as well as casework activities.

The chief reason, however, why worker scores were not higher than second year student scores seemed to be in the scoring system itself. The scoring system was based on what was satisfactory for a first year student and gave no premium for answers that were better than what was considered satisfactory for a first year student. Thus workers whose answers

might actually have been superior or creative may have been prevented from getting higher scores.[38]

In every question asking for an explanation of reasons for activity or the goal or objective sought for the activity, many of the workers would have received higher scores if there had been a "superior" category for responses. While worker responses telling what they would do or say in response to certain questions differed little from students, there were considerable differences in the reasons workers gave for their activity or the objectives or goals they sought by the activity. For students, usually only one reason or one objective was given for an activity. Workers often gave multiple explanations why an activity was necessary or a multiple statement of objectives. When multiple objectives were given they frequently included both the immediate and long term goals. For example, a student might report that the reason for an activity was to reduce the client's defensiveness; the worker might mention this but also indicate that a reduction in defensiveness was necessary to get the client to relate to worker and at the same time to help the client begin to look at his difficulties.

---

[38] Subsequent to completing analysis of the data for this study, scores were obtained on 42 welfare visitors who were administered the TASC. These scores were significantly lower than beginning first year students and had an average (mean) score of 55. Whether all of the difference between this average score and the average score for beginning first year students is due to a difference in casework knowledge cannot be determined, since the welfare visitors were judged by a different judge. The size of the difference, however, almost 20 points, would appear to indicate that the test differentiates the average welfare visitor group from beginning first year students. If this is confirmed by further study, this means the TASC can be used at admission to help in determining whether applicants are more like beginning first year students or more like the usual welfare visitor.

As a result of this finding, a supplemental scoring system has been added to the original one. This supplemental system provides four categories for scoring responses rather than the three originally used, and includes criteria for scoring certain responses as "superior." Because the reliability and validity of this supplemental system has not been tested, it has been listed separately.

Although the total scores did not discriminate between second year students and caseworkers, differences appeared when questions were divided into three levels of difficulty, "easy," "moderate," and "difficult."

Both workers and students found the same four questions significantly easier than the average question. Although there was little difference between caseworker and second year student scores on these, second year students scored higher than first year students.

Among the three most difficult questions (significantly more difficult than the average question), there were two on which students scored lower than workers. One of these was among the easiest items for workers.

The ability of the TASC to differentiate between groups who have had different amounts of education, training and experience may be exploited in a number of ways. If given to students at their entrance into school, the TASC may be used to separate them into sections according to their background knowledge of casework. It may be useful in evaluating differences in teaching methods with two otherwise matched groups. It may be useful among untrained welfare visitors in selecting groups that can benefit most from training.

All of these potential uses, however, refer to groups. The TASC should not be used to differentiate one individual's casework knowledge from another, or to determine differences in an individual's knowledge on different occasions, such as determining if an individual has learned from an educational or practice experience. Like many instruments of this type, the TASC shows only very gross differences between individuals. Any individual who is given the TASC at more than one time may make a higher or lower score on the second administration. The TASC is like any other test in this respect, that is, individual scores are subject to chance variations in responses. The standard error of an individual score is five points. This means that individual scores that are within ten points of each other are probably (95 times out of a hundred) being influenced by these chance variations. Individual scores should not be considered significantly different unless they are more than 10 points apart.

Variation in Scores:

The variation in scores from one judge to another, within any particular student or worker group was quite small. This would be expected in view of the findings about the reliability of judgments and is further evidence that judges tended to have a homogeneous viewpoint.

The amount of variation in scores within each student group was quite similar. The amount of variation within the worker group from Area 3, however, was significantly different for all judges from the variation within other worker groups. Partly this may have been related to the fact that the worker group from Area 3 was the smallest sample of all tested; partly this finding may be due to the fact that workers from Area 3 were tested on two occasions and differences in administration of the test from occasion to occasion may have occurred.

Details on the variation within each of the groups tested is shown in Tables S-1 through S-3 in the Appendix.

Validity:

At this stage in its development, the confirmation that the TASC measures casework knowledge is largely dependent on its ability to discriminate between groups at different levels of education, and on logical analysis of the questions on the test. Scores of the various groups on the TASC and their scores on other generally accepted measures of casework knowledge were sometimes similar and sometimes not. Students' TASC scores were sometimes like and sometimes unlike their casework and field instruction grades; workers' TASC scores, on the other hand, were generally found to correspond to a significant extent with ratings made by their supervisors on the scales provided them.

The finding that untrained welfare visitors, beginning first year students, ending first year students and beginning second year students, and ending second year students scored in that order on the TASC implies that the TASC is a valid measure of casework knowledge. Each of these groups would be expected to have more casework knowledge than the preceding group because each had received more training and had more opportunity to get casework knowledge. Welfare visitors tested were, in general, not planning to enter school. Because so many beginning first year students come from positions with departments of welfare, beginning first year students can be considered to be a selected sample of welfare visitors in general. Students who had finished a year could be expected to have more casework knowledge than those beginning the first year; students who had finished two years of education would be expected to know most of all.

The finding that TASC scores for some student groups correlated with casework and field instruction grades and for other student groups they did not, taken at the same time with the finding that TASC scores correlated with supervisors' ratings of workers, may mean grades lack reliability and validity. At least two other factors may influence this lack of correspondence; grades have less variation than the TASC scores; casework and field instruction grades may also measure more than the knowledge covered by the TASC. Because such a large proportion of students make "B" grades, some students who do poorly on the TASC make "B"'s and some who do well make "B"'s; because casework and field instruction grades measure such things as use of community resources, use of supervision, and use of agency policies, the correspondence of the TASC scores may be less than if both measured the same thing.

Correlations between the TASC total scores for each judge and grades for human behavior and research courses as well as correlation between the TASC scores and casework and field instruction grades for first and second year students are shown in Tables V-1 through V-6 in the Appendix. Correlations between the TASC total scores for each judge and rating scales on workers are in Tables V-7 through V-9 in the Appendix.

Comparison of Narrative and Objective Forms:

The narrative form and the objective form of the TASC were not found to be significantly different in difficulty. As might be expected, however, on the grounds that it is easier to recognize the correct answer to a question than to recall it from memory or to conceptualize it and write it in narrative form, the total responses to the objective form and some of the responses to certain questions on the objective form were slightly higher than the corresponding narrative responses.

Mean and median scores for the first year and second year students who took both the objective and narrative forms of the TASC are shown in Table 5 below.

Table 5

Comparison of Mean and Median Scores

Narrative and Objective Forms*

	Area 1 - 1st Yr. Students	Area 2 -2nd Yr.Students
Mean Score Objective	81.5	86.0
Mean Score Narrative	79.1	82.5
Median Score Objective	81.8	86.1
Median Score Narrative	78.0-83.7**	85.0-90.5

*Because the objective form lacked the second part of one question, scores on the objective form have been adjusted by adding in the average score for this missing part as taken from the narrative form.

**Median scores show ranges of medians for four professional judges.

The most difficult questions on the narrative remained the most difficult on the objective. For two of the three questions significantly more difficult than average on both forms, the objective question score was higher; for one question, the objective question score was lower. The four questions significantly easier than the average for the narrative form were also among those easiest on the objective form. Two other questions were also equally easy on the objective form although they were among the most difficult on the narrative form. These were the two diagnostic questions. Apparently students can recognize the correct diagnosis with ease but have difficulty when it is necessary that they conceptualize and write this out themselves.

Because the objective form is considered easier to score, these findings argue well for the substitution of the objective form for the narrative. The objective form, however, is considered to need further testing and is being released for experimental use only. Findings mentioned above were obtained on a total of only 78 students. Further testing is thus required to make sure that similarities between objective scores and narrative scores will continue to be found. One other reason for preferring the narrative form to the objective form, whenever one can invest the judging time required, is that the responses on the narrative form are more analogous to actual practice. In actual practice, the worker must recall what to do, conceptualize his statement, and then communicate it to the client. In actual practice he is not asked to recognize the correct statement as he is on the objective form.

Content:

Inspection of the questions on which the TASC is based shows a somewhat balanced sampling of the knowledge required for the study, diagnostic, and treatment activities of casework. Inferences based on scoring patterns, or correlations between scores for various questions for each of the various groups, however, show that groups at various stages of professional development answer the questions from different frames of references.[39]

---

[39] Intercorrelated questions were those that tended to be responded to similarly by the persons being tested. When those members of a student group who scored high on question 5, for example, also scored high on question 10, and those who scored low on question 5 also scored low on question 10, these two questions would be intercorrelated. When the high scoring students on these two questions also scored high on question 11, and the low scoring students on these two questions also scored low on question 11, then question 11 would also be intercorrelated with questions 5 and 10, and questions 5, 10, and 11 could be said to represent a pattern of responses.

Further, these inferences seem to indicate that as students gain knowledge through the educational processes, the same question may be answered in different ways.

By inspection of the wording of questions and their inferred meaning, six items appeared to measure respondents' knowledge of how to deal with the feelings necessary to complete the study process. These items required knowledge of how to reduce anxiety and discomfort that might interfere with the client's participating in the study process or asked for knowledge how to keep the study process focused on the problem at hand. Two questions were definitely diagnostic in type. One of these asked for a diagnosis on the husband in the family and the other asked for a diagnosis on the wife. Five questions could be classified as requiring knowledge of different kinds of treatment and when each kind should be offered. Three of the study items appeared to include certain aspects of treatment. These three questions required that information for the study process be obtained in such a way as to help the client gain insight into his problems.

Both the easiest and the most difficult questions were found in the treatment group of items. The most difficult questions for all groups tested were those that required explanations why certain goals were sought or why certain treatment activities were engaged in. The easiest items for all groups were also treatment items, and were those that required the respondent to permit the client to solve his own problem or to help him to do this. Diagnostic questions were among the easiest for workers; diagnostic questions, however, were among the most difficult for students.

Patterns of responses obtained were influenced by the variation among the kinds of respondents tested and the relative size of each homogeneous subgroup. When all respondents were grouped together, no pattern of responses was found. Less than 5 percent of these responses were intercorrelated with other responses. In view of the generally accepted interrelationships of various aspects of practice, this finding would suggest that the lack of patterning was associated with different frames of reference of students and workers rather than coming from any basic lack of relationships between knowledge required by the various questions.

Analysis of patterns of responses by the various groups supported this suggestion. First year students tended to answer each question as though it were separate and distinct from each other question. Evidence for this is shown by the fact that less than 10 percent of the questions were correlated significantly with each other. Workers had the most patterned responses (with about 35 percent of the questions significantly intercorrelated) and second year students fell in between with about 18 percent of the items intercorrelated. This finding probably means that responses were influenced by one or more of the following factors. Beginning students may have answered questions at least somewhat on the basis of their previous life experiences and experiences as untrained workers; the knowledge they learned in school so far may have been largely unintegrated and they may thus have seen each question as separate and unconnected. Part of the lack of interrelationship among item scores may be due to the variation in the stability or consistency with which the item was interpreted by the respondent. Part of the lack of patterning may be due to variations in the judging process. The latter influences (from

various understandings of the item or variations in judging), however, tended to be similar from group to group. Since patterns tended to develop as both students and workers obtained training and experience, the lack of interrelationship among students' responses would appear more due to lack of knowledge or lack of integration of knowledge than the questions themselves.

No patterns of responses to the various questions could be found among the few questions to which responses were correlated among students. Among workers, two patterns or clusters of correlations emerged. These patterns did not appear to be influenced by the relative difficulty of the questions so much as they appeared to be influenced by the content. One pattern showed a cluster of intercorrelated items that required knowledge of diagnosis, knowledge of how to set a treatment goal based on the diagnosis, and knowledge of an appropriate method to be used in reaching this goal. The other patterned response showed a cluster of questions all dealing with the study process and largely requiring a focusing on the immediate problems in the family situation. This did not include all questions dealing with the study activities. The lack of other patterns, and the lack of a large enough number of item intercorrelations sufficient for a factor analysis, suggests that even among the relatively homogeneous worker groups there was considerable lack of homogeneity in the frame of reference with which they responded to questions. Further analysis of responses of a more homogeneous group of respondents may throw more light on what the TASC measures.

Chapter V

SUMMARY AND FUTURE POSSIBILITIES

As indicated by the material in the previous chapter, the major goals of this study have been met. The Tulane Assessment Scale for Caseworkers has been shown to be a reliable, discriminating instrument that appears to validly measure what is commonly called social casework knowledge. Its administration and scoring time appear to be within reasonable limits.

The reliability of the TASC is on a par with or better than many instruments in common use in psychology and sociology as well as the few instruments in use in the social work field. The lack of significant differences between scores of respondents who took the test on two occasions indicates that the questions are sufficiently clear that respondents' interpretations of them are identical or similar on different occasions. The relatively high agreement between and among judges who judged responses (with a median inter-judge correlation of .77) generally shows better than satisfactory agreement among judges for an instrument of this type (which is essentially a rating scale).

The TASC scores of beginning students in their first year were significantly lower than scores of students at the end of their first year; both these two groups in turn scored significantly lower than students who had completed the second year of their Master of Social Work training. While the scoring system used in this study did not show differences between scores of practicing caseworkers with Master's degrees and second

year students who were completing their Master's work, the scoring system developed as an outcome of this study is expected to show these two groups as scoring differently.

The validity of the TASC as a measure of casework knowledge was shown by a content analysis of the questions and by its discrimination between groups at various stages of training as described above. The fact that the TASC scores sometimes corresponded to casework and field instruction grades and sometimes did not is considered to be due at least in part to lack of consistency in grading and to the fact that grades measure different things than are measured by the TASC. Support for this consideration is found in the correspondence of the TASC scores with supervisors' rating scales on the amount of casework knowledge held by practicing caseworkers with Master's degrees.

The administrative time of one and a half hours for the TASC and the fact that it can be given to groups of as many as 100 respondents at one time appears to make it a feasible test. The scoring time of fifteen minutes for the narrative form and for one minute for the objective form appears reasonable and not costly to score. The finding that clerical persons do not score responses significantly different from professional persons increases the ease of scoring and reduces the cost with which the TASC can be scored.

These findings thus appear to show that at least some of the knowledge aspects of social casework competence can be measured on the basis of what a person says or does at a particular time. The advantages of measuring casework knowledge in this way, rather than assuming a certain amount of knowledge on the basis of certain experiences which a person has undergone, are obvious.

As one looks to the future, the other aspects of competence, values, purposes, sanctions, and methods should lend themselves to similar measurements. The writer believes that measurement of these aspects of competence can be made. The methods and techniques for doing so are well known and already available. What remains is a firm decision by the profession that measurement of these aspects is desirable and the necessary commitment of time, money, and manpower to accomplish it. In fact, instruments to measure some of these aspects of competence have been developed or are under development. For example, the Oren Attitude Scale has been found to differentiate the values of social workers who were rated at various levels of competence by their supervisors or administrators. The writer is working on another instrument to measure research knowledge and attitudes to research and a second instrument that it is believed will throw some light on the thinking patterns of social workers with various degrees of competence.[40]

As these measures of other aspects of casework and social work competence are worked on, further work can also be done on the TASC. Mention is made of some of the other studies that can be made with this instrument in the hope that other researchers, doctoral students seeking dissertation topics, and agency personnel looking for answers to some of the questions mentioned in the introduction will be stimulated to carry out some of these studies.

---

[40] The Oren Attitude Scale has not been published but may be obtained from Ann Oren at the University of Minnesota School of Social Work. The first of the writer's instruments referred to is called "Measurements of Attitudes and Research Knowledge," and is referred to as the MARK. This instrument has been under development for three years and is almost ready for release. The second of the writer's instruments is called "Conceptual Similarities" and is now in the beginning stage.

More information can be obtained about both the upper and lower range of scores which are possible on the TASC. Further testing of the instrument is needed using the "superior" category of scoring that has been added to the three categories, "satisfactory, innocuous, and unsatisfactory" used in this study. This test could be conducted on a homogeneous group of workers with from three to five years experience or a homogeneous group of agency supervisors who are considered outstanding in the field. Scores of the latter group would help to determine the upper limits of scoring that are possible on this instrument. More information can be obtained about the lower range of scores on the TASC. The instrument can be tested further with welfare visitors, case aides, and volunteers in the field of social work to determine whether it will discriminate these from beginning first year students. Its use on college students and on samples of the general population might also indicate not only the range of scores for these groups but something about their characteristic responses to social work problems and how they differ from professional groups.

Use of the TASC on the same groups in a before and after fashion will help to obtain information on how well it can discriminate the same group at different times. A number of tests like this are necessary to make sure that lack of changes found in a before and after study are not due to lack of sensitivity of the instrument.

Considerable testing of the objective form can be carried out. It too can be tested on a group of superior workers or supervisors, and on welfare visitors, case aides, and volunteers to determine its comparison to the narrative form. The objective form may also be given to groups

who have not seen the motion picture to determine what kind of score can be made simply on the basis of logical analysis of the questions and answers.

Qualitative analysis of responses to various questions may yield information on the knowledge base held by workers and on characteristic patterns of response. For example, there is a question whether students tend to make more exploratory type responses in comparison with workers who might be found to make more directive types of responses.

Analysis of the responses to various questions may also help to discover which tasks can be done adequately by welfare visitors, case aides and social work technicians and which should be left to personnel with Master's degrees. If the scores on certain questions made by welfare visitors, case aides and social work technicians do not differ from those made by persons with Master's degrees this would provide some evidence that this kind of person could do this task as well as the persons with the Master's degrees.

In using the TASC for these goals or in ways already tested in this particular study certain cautions should be followed. The vicarious nature of the practice experience used in the TASC always leaves unanswered the question whether the respondent can do what he says should be done. It is at least a beginning, however, in measuring what one can do, rather than inferring this from the education and from experiences the person has undergone. Caution should also be used in interpreting the meaning of individual TASC scores. While it will discriminate one group from another, it should not be used to discriminate one individual from another.

In time, as casework practice changes, and as the knowledge required by practice differs from the present, the responses considered satisfactory and superior for this particular instrument will probably require a change. It may also be found that the situations shown in the motion picture are no longer representative of practice; another motion picture may then be used. At any rate, the method demonstrated by this particular study can be applied to future vicarious practice experiences, and the knowledge gained in developing this test can be used to reduce the development time of another instrument.

APPENDIX

Experience of Respondents

Table E-1

Average Experience of Caseworkers Before and After Completing MSW, by Area

	Before Education		After MSW	
Area	Median	Mean	Median	Mean
1	1	2.5	7	7.0
2	0	0.0	3	2.8
3	0	1.9	4	4.2
All	0	1.4	4	5.0

Table E-2

Range of Experience of Caseworkers Before and After Completing MSW*

	Before any Education				After MSW			
Experience	Total	Area 1	Area 2	Area 3	Total	Area 1	Area 2	Area 3
15+	0	0	0	0	9	4	0	5
12-14	1	1	0	0	7	6	0	1
9-11	2	1	0	1	24	5	1	18
6-8	11	8	1	2	34	10	13	11
3-5	9	3	0	6	25	15	8	2
2	13	9	0	4	9	3	6	0
1	13	6	0	7	2	2	0	0
0	61	17	27	17	0	0	0	0
Total	110	45	28	37	110	45	28	37

* In computing means, experience over 15 years was counted as 15; area 1, however, had workers with as much as 28 years of experience.

Table E-3

Average Experience of Second Year Students Before & Between Two Years of MSW Education by Area

(Both Students at the Beginning and Those at the End of Their Second Year are Included)

	Before Education		Between the Two Years	
Area	Median	Mean	Median	Mean
1	1.0	1.4	0.0	0.9
2	1.0	2.0	1.0	2.1
3	1.0	2.5	1.0	1.9
All	1.0	1.9	1.5	1.6

Table E-4

Range of Experience of Second Year Students Before and Between Two Years of MSW Education by Area

(Both Students at the Beginning and Those at the End of Their Second Year are Included)

	Before Education				Between Two Years			
Experience	Total	Area 1	Area 2	Area 3	Total	Area 1	Area 2	Area 3
15+	1	0	0	1	1	0	0	1
12-14	1	0	1	0	0	0	1	0
9-11	2	0	0	2	1	0	0	1
6-8	6	2	3	1	5	0	3	2
3-5	17	5	5	7	15	4	7	4
2	20	7	9	4	14	5	7	2
1	24	5	13	6	22	7	12	3
0	45	18	12	15	58	21	13	24
Total	116	37	43	36	116	37	43	36

Table E-5

Average Experience of First Year Students Before Entering School, by Area

(Both Students at the Beginning and End of Their First Year are Included)

	Median Experience	Mean Experience
Area 1	1.0	1.6
2	1.0	2.0
3	0.0	1.7
All	1.0	1.8

Table E-6

Range of Experience of First Year Students Before Entering School by Area

(Both Students at the Beginning and End of Their First Year are Included)

Experience	Area 1	Area 2	Area 3
12-14	0	1	1
9-11	1	3	0
6-8	2	1	2
3-5	10	6	6
2	12	6	5
1	14	12	6
0	24	20	21
Total	63	49	40

Table E-7

Percent of Each Group Tested Who Saw the Motion Picture Before Being Tested

	Workers	Second Yr. Students	1st Yr. Students
Area 1	20%	13%	10%
Area 2	28%	14%	6%
Area 3	24%	10%	10%
All	24%	13%	8%

Table E-8

Percent Tested Who Had Discussed Motion Picture Before Being Tested

	Workers	Second Yr. Students	1st. Yr. Students
Area 1	9%	11%	3%
Area 2	7%	None	2%
Area 3	11%	2%	2%
All	9%	4%	3%

TASC First Year Students

(Both Students at Beginning and End of 1st Year are Included)

Relation Between Background Factors and Total Scores on TASC

Table E-9 Area 1*

	Exper. Before School	Saw Picture**	Disc. Picture**
Judge			
J1	-.08	NA	NA
J2	-.01	NA	NA
J3	-.04	NA	NA
J4	-.07	NA	NA
J5	-.09	NA	NA
J6	-.11	NA	NA

* Correlation necessary for significance at 5% level = .25
** No correlation computed; only 2 students saw picture previously; only one discussed picture previously.

Table E-10 Area 2*

	Exper. Before School	Saw Picture	Disc. Picture
Judge			
J1	.07	NA**	NA**
J2	.12	NA	NA
J3	.04	NA	NA
J4	.20	NA	NA

* Correlation necessary for significance at 5% level = .30
** Correlations not computed; only 4 students saw picture; only 1 discussed picture.

Table E-11 Area 3*

	Exper. Before School	Saw Picture	Disc. Picture
Judge			
J1	-.24	NA**	NA**
J2	-.16	NA	NA
J3	-.17	NA	NA
J4	-.14	NA	NA

*Correlation necessary for significance at 5% level = .27
**Only 3 students saw and only 1 discussed picture; no correlations computed.

TASC Second Year Students

Relation Between Background Factors and Total Scores on TASC

Table E-12 Area 1*

	Exper. Before School	Exper. Between 2 yrs.	Saw Picture	Disc. Picture
J1	-.43	.07	-.02	.03
J2	-.31	.04	-.12	-.11
J3	-.45	-.02	.02	.01
J4	-.50	.19	.03	.19

*Correlation necessary for significance at 5% level = .32

Table E-13 Area 2*

	Exper. Before School	Exper. Between 2 yrs.	Saw Picture	Disc. Picture
J1	-.06	.11	NA**	NA**
J2	-.25	-.06	NA	NA
J3	-.18	-.15	NA	NA
J4	-.29	-.10	NA	NA

*Correlation necessary for significance at 5% level = .30
**No correlations computed; only 4 saw and 1 discussed picture.

Table E-14 Area 3*

	Exper. Before School	Exper. Between 2 yrs.	Saw Picture	Disc. Picture
J1	.00	.02	NA**	NA**
J2	.00	.05	NA	NA
J3	-.05	.01	NA	NA
J4	.11	.07	NA	NA
J5	.08	.00	NA	NA
J6	.10	.00	NA	NA

*Correlation necessary for significance at 5% level = .32
**No correlations computed; only 6 saw and 1 discussed picture.

TASC Workers

Relation Between Background Factors and Total Score on the TASC

Table E-15 Area 1*

	Exper. Before School	Exper. Between 2 yrs.	Exper. After MSW	Saw** Picture	Disc.** Picture
J1	-.05	-.19	-.09	.19	.41
J2	-.16	-.16	-.25	.39	.32
J3	-.12	-.12	-.11	.27	.30
J4	-.00	-.26	-.04	.38	.39

*Correlation necessary for significance at 5% level = .29
**Based on 20% seeing picture and 10% discussing it.

Table E-16 Area 2*

	Exper. Before School	Exper. Between 2 yrs.	Exper. After MSW	Saw** Picture	Disc.** Picture
J1	-.02	-.06	.06	-.10	.02
J2	.14	.16	.07	.05	.03
J3	-.01	.20	-.17	.01	.06
J4	.06	.19	-.01	.08	.10
J5	.34	.15	.00	.07	.08
J6	.13	.19	-.05	-.10	.03

*Correlation necessary for significance at 5% level = .30
**Based on approximately 24% who saw motion picture and 10% who discussed it.

Table E-17 Area 3*

	Exper. Before School	Exper. Between 2 yrs.	Exper. After MSW	Saw*** Picture	Disc.*** Picture
J1	NA**	NA**	-.29	.31	.46
J2	NA	NA	-.30	.43	.62
J3	NA	NA	-.30	.55	.58
J4	NA	NA	-.36	.39	.53

*Correlations necessary for significance at 5% level = .37
** These workers did not answer these items.
*** Because only 30% saw picture and 10% discussed it these correlations are probably inflated.

Reliability of Responses*
Test-Retest

Table R-1 - First Year Students Area 1

Judge	RJ1	RJ2	RJ3	RJ4
J1	.59	.50	.61	.70
J2	.61	.68	.65	.67
J3	.57	.64	.62	.65
J4	.61	.62	.63	.75

Table R-2 - Second Year Students Area 3

Judge	RJ1	RJ2	RJ3	RJ4
J1	.26	.08	.22	.21
J2	.38	.26	.39	.29
J3	.39	.31	.46	.45
J4	.52	.36	.49	.44

*The reason for these rather wide differences in test-retest correlations is not clear. It may represent actual "between group" variation. More likely, however, in this case, it represents differences in test administration. The group from area one were all volunteers; the entire group from area 3 was retested. The group from area one were tested early in the morning when they were fresh; the group from area three were tested after a hard day in classes.

TASC First Year Students

Interjudge Reliability Based on Total Scores, on Narrative Form of the TASC

Table R-3 Area 1

	J1	J2	J3	J4	J5	J6
J1	--	.74	.73	.74	.65	.59
J2		--	.75	.77	.62	.66
J3			--	.79	.83	.70
J4				--	.70	.66
J5					--	.77
J6						--

Table R-4 Area 2

	J1	J2	J3	J4
J1	--	.64	.54	.63
J2		--	.78	.76
J3			--	.70
J4				--

Table R-5 Area 3

	J1	J2	J3	J4
J1	--	.64	.69	.67
J2		--	.73	.76
J3			--	.76
J4				--

TASC Second Year Students

Interjudge Reliability Based on Total Scores on Narrative Form of the TASC

Table R-6 Area 1

	J1	J2	J3	J4
J1	--	.74	.79	.72
J2		--	.81	.77
J3			--	.82
J4				--

Table R-7 Area 2

	J1	J2	J3	J4
J1	--	.69	.63	.65
J2		--	.71	.63
J3			--	.73
J4				--

Table R-8 Area 3

	J1	J2	J3	J4	J5	J6
J1	--	.77	.71	.75	.72	.76
J2		--	.77	.78	.70	.76
J3			--	.79	.76	.82
J4				--	.83	.89
J5					--	.85
J6						--

TASC Workers

Interjudge Reliability Based on Total Scores on Narrative Form of the TASC

Table R-9 Area 1

	J1	J2	J3	J4
J1	--	.65	.76	.79
J2		--	.63	.65
J3			--	.67
J4				--

Table R-10 Area 2

	J1	J2	J3	J4	J5	J6
J1	--	.56	.50	.71	.48	.70
J2		--	.50	.72	.63	.63
J3			--	.77	.56	.60
J4				--	.55	.65
J5					--	.74
J6						--

Table R-11 Area 3

	J1	J2	J3	J4
J1	--	.87	.85	.84
J2		--	.88	.91
J3			--	.91
J4				--

Interjudge Reliability on Each Item

Table R-12

Item No.	J1 & 2	J3 & 4	J5 & 6
1	.79	.67	.90
2A	.51	.70	.77
2B	.43	.28	.62
3A	.27	.43	.72
3B	.14	.14	.19
4A	.58	.20	.72
4B	.40	.31	.44
5	.49	.55	.57
6	.28	.58	.47
7A	.56	.63	.35
7B	.50	.51	.46
8	.39	.49	.44
9A	.99	.99	.98
9B	.92	.95	.84
10A	.41	.70	.51
10B	.38	.51	.16
11A	.33	.11	.50
11B	.47	.17	.57
12A	.37	.14	.64
12B	.47	.17	.56

Table D -1

Mean Scores for All Groups by Stage of Professional Development and by Judge

	Judge 1	Judge 2	Judge 3	Judge 4	Average for four Prof. judges	Judge 5*	Judge 6*
Beginning 1st.Yr.							
Area 3	71.2	78.5	81.4	68.7	74.9		
Ending 1st Yr. & Beginning 2nd Yr.							
Area 1	78.2	80.8	80.5	76.9	79.1	77.4	75.8
Area 1 Retest	80.5	85.6	84.8	78.0	82.5		
Area 2	83.6	86.4	84.8	79.5	83.2		
Area 3	75.0	80.8	84.8	73.1	78.4	78.0	76.2
Area 3 Retest	72.8	79.8	83.7	74.0	77.6		
Ending 2nd Yr.							
Area 1	88.6	86.2	85.8	81.8	85.6		
Area 2	84.2	87.8	80.3	81.5	83.4		
Caseworkers							
Area 1	79.6	85.4	86.9	77.7	82.4		
Area 2	87.4	91.6	88.6	81.5	87.6	86.0	87.2
Area 3	82.6	85.5	87.8	77.0	84.5		

* Judges 5 and 6 were the clerical judges.

Table D-2

Median Scores for All Groups by Stage of Professional Development and by Judge

	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5*	Judge 6*
Beginning 1st Yr.						
Area 3	73.5	79.0	74.4	68.3		
Ending 1st Yr. & Beginning 2nd Yr.						
Area 1	81.2	83.7	80.5	78.0	80.3	81.5
Area 1 Retest	84.1	85.4	86.1	79.8		
Area 2	84.5	80.0	85.6	80.6		
Area 3	76.0	82.0	85.7	84.0	77.8	78.3
Area 3 Retest	76.8	75.8	85.8	79.6		
Ending 2nd Yr.						
Area 1	84.0	85.8	84.0	78.3		
Area 2	85.0	88.0	87.7	90.5		
Caseworkers						
Area 1	80.9	88.0	90.3	78.9		
Area 2	85.7	91.0	89.1	81.0	84.7	83.8
Area 3	81.7	85.8	90.5	78.8		

*Judges 5 and 6 were the clerical judges.

Standard Deviation by Stage of Development and by Judge for Each Group of Respondents

Table S-1

First Year Students

Judge	Area 1	Area 2	Area 3
J1	11.46	10.54	8.15
J2	11.31	8.51	9.57
J3	10.22	9.04	9.64
J4	9.64	9.23	10.67
	N=63	N=40	N=49

Table S-2
Second Year Students

J1	11.64	10.28	10.68
J2	10.75	9.05	9.75
J3	10.19	9.81	8.70
J4	11.01	9.00	11.87
	N=37	N=36	N=43

Table S-3

Caseworkers

J1	10.81	8.12	16.25
J2	9.82	7.89	15.63
J3	10.77	9.60	17.26
J4	10.55	10.47	16.32
	N=45	N=37	N=28

TASC First Year Students

Validity Correlations Between Each Judge's Total Score and Course Grades

Correlations significant at the 5% level are starred "*"

Table V-1 - area 1

Judge	CW II	CW IV	F.Ins.II	F.Ins.IV	Hum.Beh.I	Hum.Beh.II	Res.I	Res.II
J1	-.12	--	.05	--	-.21	--	.16	--
J2	-.14	--	.07	--	-.24	--	.19	--
J3	.02	--	.05	--	-.11	--	.27*	--
J4	.03	--	.16	--	-.12	--	.26*	--
J5	.08	--	.12	--	-.14	--	.22	--
J6	.10	--	.14	--	-.18	--	.29*	--

Table V-2 - area 2

Judge	CW II	CW IV	F.Ins.II	F.Ins.Iv	Hum.Beh.I	Hum.Beh.II	Res.I	Res.II
J1	.04	--	.16	--	.08	--	.02	--
J2	.06	--	.19	--	.09	--	.21	--
J3	-.16	--	.18	--	.03	--	.26	--
J4	.00	--	.00	--	.01	--	.15	--

Table V-3 - area 3

Judge	CW I	CW II	F.Ins. II	Hum.Beh.I	Hum.Beh.II	Res. I
J1	.20	-.06	.18	.04	-.21	.10
J2	.06	-.04	.06	-.03	-.18	-.17
J3	.07	.07	.00	-.10	-.13	.20
J4	.04	-.14	.14	-.18	-.33*	.11

TASC Second Year Students

Validity Correlations Between Total Score for Each Judge and Grades in Various Courses

Correlations significant at the 5% level are starred "*"

Table V-4 area 1

Judge	CW II	CW IV	F.Ins. II	F.Ins IV	Path I	Path II	Res. I	Res. II
J1	-.10	-.12	-.13	.02	-.33*	-.04	.01	.15
J2	-.03	-.02	.05	.12	-.13	.15	-.08	.12
J3	.10	-.03	-.18	.03	-.23	.06	.11	.22
J4	-.03	-.13	-.09	.06	-.22	.05	-.00	.19

Table V-5 area 2

Judge	CW II	CW IV	F.Ins. II	F.Ins.IV	Path I	Path II	Res. I	Res.II
J1	-.15	-.13	-.08	-.25	.25	.06	-.12	-.07
J2	-.28	-.14	-.29	-.37*	.34*	-.07	.01	-.02
J3	-.18	.01	.01	-.24	.32*	-.20	.27	.15
J4	-.06	-.08	-.12	-.13	.30	-.19	.17	-.01

Table V-6 area 3

Judge	CWII	CW IV	F.Ins. II	F.Ins.IV	Path I	Path II	Res.I	Res.II
J1	.29*	--	.10	.14	.34*	--	.35*	--
J2	.40*	--	.25	.28*	.30	--	.21	--
J3	.29*	--	.10	.16	.31*	--	.20	--
J4	.40*	--	.30*	.00	.32*	--	.33	--
J1R	.31*	--	.37*	.42*	.14	--	.27	--
J2R	.15	--	.32*	.13	-.15	--	.33*	--
J3R	.35*	--	.43*	.30*	.16	--	.29*	--
J4R	.40*	--	.49*	.22	.12	--	.32*	--
J5	.38*	--	.23	.18	.30*	--	.36*	--
J6	.29*	--	.20	.34*	.39*	--	.29*	--

TASC Workers

Validity Correlations between each Judge and Supervisor's Rating of Workers Knowledge

Correlations significant at the 5% level are starred "*"

Table V-7 area 1

	Knowledge Held	Knowledge Used
J1	.33*	.21
J2	.50*	.47*
J3	.36*	.25
J4	.33*	.26

Table V-8 area 2

	Knowledge Held	Knowledge Used
J1	.05	.12
J2	.00	.01
J3	.30*	.07
J4	.15	.02
J5	.23	.04
J6	.28	.01

Table V-9 area 3

	Knowledge Held	Knowledge Used
J1	.27	.24
J2	.29	.33
J3	.21	.21
J4	.21	.24

Table N-O-1

Correlation of Narrative and Objective Forms

Judge	Area 1 - 1st yr. students	Area 2 - 2nd year students
J1	.61	.16
J2	.69	.25
J3	.66	.26
J4	.66	.21

BIBLIOGRAPHY

Books

Brieland, Donald. "An Experimental Study of the Selection of Adoptive Parents at Intake," Child Welfare League of America, New York, 1959.

Flexner, Abraham. Is Social Work A Profession? "Proceedings of the National Council of Charities and Corrections." Chicago: Hildman Printing Co., 1919, pp. 576-90.

Hollis, Florence. A Study of the Case Work Performances of Graduates as a Measure of the Effectiveness of Professional Training. Family Service Association of America, 1941.

Kaufman, Margaret, and Malmud, Helen. Evaluations of Staff Members in Private Family Agencies. Family Service Association of America, 1940.

LeBeaux, Charles, and Wilensky, Harold. Industrial Society and Social Welfare. 1958. Russell Sage Foundation.

Thomas, E. J. and McLeod, Donna, et. al., "In Service Training and Reduced Workload, Experiment in a State Department of Welfare," N.Y., Russell Sage 1960.

Whitehead, A. N. Aims of Professional Education. New York: 1929

Articles

Ballard, Robert G., and Mudd, Emily H. "Some Theoretical and Practical Problems in Evaluating Effectiveness of Counseling," Social Casework, Vol. 38 (1957), 533-538.

Bartlett, Harriett M. "Toward Clarification and Improvement of Social Work Practice," Social Work, Vol. 3 (April, 1958), pp. 3-9.

Bruck, Max. "The Relationship Between Student Anxiety, Self-Awareness, and Self-Concept and Student Competence in Casework," Social Casework, Vol. 44 (1963), 125-31.

Chichester, E. Elizabeth, et al. "Field Work Criteria For Second Year Casework Students," Principles and Techniques in Social Casework, 1940-1950. Edited by Cora Casius, Family Service Association of America, (1953), pp. 233-246.

Dowling, Jessie P. "Stages in the Progress of First Year Students in the Veterans Administration," Social Casework, (Jan. 1952),pp.13-18.

Giedt, Harold. "Comparison of Visual, Content and Auditory Cues in Interviewing." Journal of Consulting Psychology, Vol. 19, No. 6, 1955, pp. 407-416.

Graham, Henry M., and Blumenthal, David. "Why We Failed - As Clients See It," Family Service Highlights, Vol. 16 (June, 1955) 91-94.

Gordon, William E. "A Critique of the Working Definition," Social Work, October 1962, pp. 4-13.

Hanford, Jeanette. "Standards of Measurement Used in Staff Evaluation in a Private Family Agency," National Conference of Social Work, 1952.

Hendricks, Thomasine, "Social Work Performance Standards and Evaluation," Social Casework, Vol. 39 (1958), 391-96.

Heyman, Margaret M. "A Study of Effective Use of Social Workers in a Hospital. Selected Findings and Conclusions," Social Service Review, Vol. 35 (1962), 414-29.

Kadushin, Alfred. "A Comparison: Dutch and American Expectations Regarding Behavior of the Caseworker," Social Casework, Vol. 41 (1960), 503-11.

Kadushin, Alfred. "Testing Diagnostic Competence - A Problem for Research," (April 1961 - mimeod).

Lewis, Mary, et al. "An Experimental Design for First Year Field Instruction," Tulane Studies in Social Welfare IV, 1962.

McCaffery, Miriam. "Criteria for Student Progress in Field Work," Journal of Case Work, (Jan., 1947).

Miller, Roger. "An Experimental Study of the Observational Process in Casework," Social Work , 4, 1958. 96-102.

Preston, Malcolm; Mudd, Emily H.; and Froscher, Hazel B. "An Experimental Study of a Method for Abstracting Content of Social Case Records," Journal of Abnormal and Social Psychology, Vol. V, 1950. 628-646.

Regensburg, Jeannette. "Some Thoughts on Being a Professional Social Worker," Social Casework, Vol. 40 (1959), 220-25.

Richmond, Mary. "The Long View, Papers and Addresses by Mary Richmond," ed. by Joanna Calcord. Russell Sage Foundation. (1930) p. 95.

Schubert, Margaret. "Field Work Performance: Achievement Levels of First-Year Students in Selected Aspects of Casework Service," Social Service Review, Vol. 32 (1958), 120-137.

Schubert, Margaret, "Field Work Performance: Suggested Criteria for Grading," Social Service Review, Vol. 32 (1958), 247-57.

Shyne, Ann W. "Evaluation of Results in Social Work," Social Work, Vol. 8 (October, 1963), 26-38.

Strupp, Hans. "The Effect of the Psychotherapist's Personal Analysis Upon His Techniques," Journal of Consulting Psychology, 19, 1955, pp. 197-204.

Thomas, Edwin J.; McLeod, D. L.; and Hylton, L. F. "The Experimental Interview: A Technique for Studying Casework Performance," Social Work, Vol. 5 (July, 1960), 52-58.

Turner, Virginia. "The Happy Congruence of Theoretical and Practical Considerations in Instrument Validation" paper presented at the National Conference of Social Welfare, Los Angeles, California, May, 1964.

Varley, Barbara K. "Socialization in Social Work Education," Social Work, Vol. 8 (July, 1963), 102-109.

Walsh, Margaret A. "Supervisory Appraisal of the Second-Year Student in Field Work," Social Casework, Vol. 41 (1960), 530-33.

Miscellaneous

Carr, Lewis W. "The Relationship Between Use of Knowledge in Practice and Effectiveness of Practice as Seen in the Development of Psycho-social Diagnostic Impressions and Prediction by Social Caseworkers." Doctoral dissertation, Washington University, St. Louis, Missouri, 1961.

Conner, Geraldine L. "A Study of Indices of Competence Used by Supervisors to Evaluate Casework Performance." Doctoral dissertation, Washington University, St. Louis, Missouri, June, 1962.

Piven, Herman. "Professionalism and Organization Structure: Training and Agency Variables in Relation to Practitioner Orientation and Practice." Doctoral dissertation, Columbia University School of Social Work, New York, N.Y., 1961.

Schubert, Margaret S. "The Quality of the Student's Casework Performance: A Study of Field Work Students at the End of the First Year in a Public Assistance Setting." Doctoral dissertation, School of Social Work, Chicago University, 1956.

Schwartz, Edward E. "Work Measurement for Performance Budgeting in a Child-Placing Agency." Doctoral dissertation, School of Social Work, Chicago University, 1955.

Shively, Frances. "A Study of the Casework Relationship" Doctoral dissertation, George Warren Brown School of Social Work, Washington University, St. Louis, Missouri (1962).

American Association of Schools of Social Work. Testing the Product of the Schools by the Performance of the Graduates. Workshop Report, 31st Annual Meeting. American Association of Schools of Social Work. January, 1950.

National Association of Social Workers. Building Social Work Knowledge. A Report of a Conference. National Association of Social Workers. 1964.

Council of Social Work Education, Official Curriculum Policy Statement. 10/19/62 (mimeod)

Committee on Study of Competence, National Association of Social Work 11/1/64 (mimeod)